Psych[illegible] Bul[illegible]

[illegible], Editor

[illegible]NTS

[illegible], No. 5

The Psychological Bulletin contains evaluative reviews of research literature and articles on research methodology in psychology. This JOURNAL *does not publish reports of original research or original theoretical articles.*

Manuscripts should be sent to the Editor, Harry Helson, Department of Psychology, University of Texas, Austin, Texas.

Preparation of articles for publication. Authors are strongly advised to follow the general directions given in the *Publication Manual of the American Psychological Association* (1957 Revision). Special attention should be given to the section on the preparation of the references (pp. 50–60), since this is a particular source of difficulty in long reviews of research literature. *All copy must be double spaced, including the references.* All manuscripts should be submitted *in duplicate*, one of which should be an original typed copy; author's name should appear only on title page. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail and to check carefully the typing of the final copy.

Reprints. Fifty free reprints are given to contributors of articles and notes. Authors of early publication articles receive no gratis reprints.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Publication Manager

BEATRICE B. HAMLIN
Editorial Assistant

Communications—including subscriptions, orders of back issues, and changes of address—should be addressed to the American Psychological Association, 1333 Sixteenth Street N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $8.00 (Foreign $8.50). Single copies, $1.50.

PUBLISHED BIMONTHLY BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Menasha, Wisconsin
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Entered as second class matter at Washington, D.C., and Menasha, Wisconsin. Second-class postage paid at Washington, D.C., and at additional mailing offices. Printed in U.S.A.

VOL. 56, NO. 5 SEPTEMBER, 1959

Psychological Bulletin

TECHNIQUES FOR THE STUDY OF GROUP STRUCTURE AND BEHAVIOR: I. ANALYSIS OF STRUCTURE[1]

MURRAY GLANZER[2] AND ROBERT GLASER
American Institute for Research and University of Pittsburgh

The main purpose of this paper is to examine the work on the structure of groups or teams. The emphasis will be on communication structure, an aspect of group behavior that has received much attention in the recent experimental and theoretical literature. The term "structure" refers here to a relationship in a group, e.g., "communicates to."

The following questions initiated this survey:

1. How can the interactions or communications of a group, its structural characteristics, be measured?
2. How are structural characteristics related to group performance?

Several areas contribute answers. The areas include sociometry, the mathematical techniques growing out of sociometry, and the group network studies. The contribution of mathematical techniques in answering the first question will be considered here. Some answers to the second question will be considered in a subsequent paper.

SOCIOMETRIC AND RELATED TECHNIQUES

There has been a considerable development of techniques for the description and analysis of the relationships between group members. The initial stimulus for this work came from the area of sociometry. Two major changes have, however, occurred in these techniques since their inception.

1. Originally, they were concerned solely with the pattern or structure of likes and dislikes within a group. They soon developed, however, to include any pattern of relation.
2. Originally, they were primarily graphical. Now they include the use of mathematical techniques.

Examples of the original approach can be found in the work of Moreno (1934) and Jennings (1950). The sociogram or graphic presentation of the relationships in a group will not be considered here. Moreover, since many of the indices developed out of earlier sociometric work are closely tied to the use of the sociogram, and also because they have been summarized in other surveys (Lindzey & Borgatta, 1954; Proctor & Loomis, 1951), the emphasis here will be on

[1] Prepared under Contract N7 onr-37008, NR-154-079, between the American Institute for Research and the Office of Naval Research, Psychological Sciences Division, Personnel and Training Branch, as part of a research project on team training and performance. The authors wish to thank Ardie Lubin and R. Duncan Luce for their helpful reviews of the paper.

[2] Now at Walter Reed Army Institute of Research.

the mathematical approaches. The survey by Proctor and Loomis (1951) gives a full treatment of the work on indices and statistical analysis that stems directly from the sociometric tradition.

Work on the techniques developing out of sociometry can be categorized into the following areas: (*a*) construction of indices for group and individual characteristics; (*b*) enumeration of structures; (*c*) comparison of groups; (*d*) analysis of subgroups; (*e*) assignment of individuals to subgroups; (*f*) other approaches: graph theory, logic of relations. Although in most cases the techniques were devised for the relationships of choice and rejection, they can be applied directly to relationships such as "interacts with" or "communicates with." One of the aims of this survey is to set forth the available techniques and to indicate their possible application. Most of the techniques discussed here have not been applied extensively. Their practical and theoretical usefulness cannot, therefore, be definitively evaluated at this point.

The first major step in the mathematical treatment of sociometric material was noting that it could be cast in the form of matrices, and the operations of matrix algebra applied to it. This step was made by Forsyth and Katz (1946). Group structure is summarized by a matrix with cells, a_{ij}. The subscripts i and j identify the row and column, respectively, of the cell entry. If a given relationship exists between individuals i and j (e.g., i speaks to j) then $a_{ij} \neq 0$. If not, then $a_{ij} = 0$. The treatment of the diagonal, a_{ii}, depends on the purposes of the investigator. In most cases, zeroes will be entered in the diagonal.

The simplest type of matrix to summarize the relations within a group is one that simply notes whether or not the given relationship exists between a pair of individuals. In this case $a_{ij} = 1$ or 0. For example, if in a four-man group it is found that *a* speaks to (or likes) *c*, *b* does the same to *a*, *c*, and *d*; *c* to *b* and *d*; and *d* to *a* and *b*; the matrix would appear as in Table 1.

TABLE 1

MATRIX WITH BINARY ENTRIES

		Receiver			
		a	*b*	*c*	*d*
Sender	*a*	—	0	1	0
	b	1	—	1	1
	c	0	1	—	1
	d	1	1	0	—

In some cases, the investigator is dealing with a relationship that varies in strength or frequency. For example, it may be that *b* speaks frequently to *a*, infrequently to *c*, and frequently to *d*, etc. In this case, the matrix might appear as in Table 2. The positive cell entries have been

TABLE 2

MATRIX WITH WEIGHTED ENTRIES

	a	*b*	*c*	*d*
a	—	0	1	0
b	2	—	1	2
c	0	1	—	2
d	2	2	0	—

weighted according to frequency. Here $a_{ij} > 0$ if the relation exists between i and j.

In other cases, the investigator is dealing with a bipolar relationship that can be placed on a continuum (e.g., strongly rejects, rejects, does not respond to, chooses, strongly chooses). In such a case, $a_{ij} \neq 0$ if individual i responds to j. The sign of the entry may then be used to indi-

cate whether the individual accepts (positive) or rejects (negative) another individual. The absolute size of the entry is used to indicate strength of the response.

Suppose, for example, that in a three-man group *a* rejects *c* and chooses *b; b* strongly chooses *c* and *c* strongly rejects both *a* and *b*. The matrix would then appear as in Table 3.

TABLE 3

MATRIX WITH WEIGHTED BIPOLAR ENTRIES

	a	*b*	*c*
a	—	1	−1
b	0	—	2
c	−2	−2	—

Katz (1947) has indicated some of the possibilities opened up by the use of matrices. These include techniques for writing equations describing over-all changes in group structure. As will be seen below, matrices also permit many types of complex analysis by relatively simple operations.

CONSTRUCTION OF INDICES FOR GROUP AND INDIVIDUAL CHARACTERISTICS

Sociometry has been prolific in the construction of indices. The meaning of these indices is usually fairly clear and their computation relatively simple. A typical index is one cited (Proctor & Loomis 1951) as a measure of group cohesion. The index is the number of mutually chosen pairs divided by $n!/(n-2)!\,2!$, the number of ways that a pair of individuals can be drawn from a group of n individuals. Since indices of this type have been presented in detail by Proctor and Loomis (1951), only those not discussed elsewhere are considered here.

Group Indices

Several indices of group characteristics have been suggested for the extent to which the group is centered on a small number of individuals. These indices are based on the variance of the column sums of the choice matrix. One of these is the index of concentration suggested by Katz (1954).

Hohn (1953) has developed an index based on the ratio between obtained variance of column sums (s) and the maximum possible variance for the special case of the matrix with weighted entries in which each individual ranks all other members of the group in order of preference. The index is called the hierarchy index with the following formula.

$$h=\frac{12}{n(n^2-1)(n-2)^2}\left(\sum_i s_i^2-\frac{n^3(n-1)^2}{4}\right) \quad [1]$$

Landau (1951a) earlier developed a similar hierarchy index for another special case, that of dominance relations.

Individual Indices

Much sociometric work is concerned with specifying characteristics of the individual in the group. For example, the total number of choices received and made by an individual is used to indicate his "popularity" and his "outgoingness" respectively. If Table 1 were a matrix of communications, *b* would be the most productive of output and receive the same amount of input as the other members.

The simplest technique in describing the individual is to use the sum of the rows and columns as above. This technique, however, makes it difficult

to compare individuals in groups of different sizes (Criswell, 1950; Edwards, 1948). Various refinements have therefore been introduced; such as, weighting the sum for each individual by the maximum number it is possible for him to send or receive. Another difficulty with the use of a simple row or column index is that it does not take into account indirect as well as direct connections between individuals. There may be an important difference between the individual who is chosen by three group members who are themselves chosen by many others and the individual who is chosen by three "isolates."

Status Index

Katz (1953) has developed a status index that takes account of such indirect links. He makes use of the fact that in a matrix with binary entries, as in Table 1, the powers of the matrix give the number of indirect connections to each member of the group. Thus, if the matrix in Table 1 has zeroes placed in the diagonal and is squared, that is, multiplied by itself, the matrix in Table 4 is obtained. The entries in the squared matrix indicate the number of two-link connections between each pair of group members. For example, there is one two-link connection between a and b ($a \rightarrow c \rightarrow b$), none between a and c, and two between c and a ($c \rightarrow b \rightarrow a$ and $c \rightarrow d \rightarrow a$). Cubing the original matrix would give a matrix all of whose entries are positive, indicating that every member of the group has one or more three-link connections with every other group member.

TABLE 4

MATRIX OF TWO-LINK CONNECTIONS

0	1	0	1
1	2	1	1
2	1	1	1
1	0	2	1

The status index, T, for each individual may be computed as the total of all direct and indirect links to the individual. These may be obtained from the column sums of the original matrix, plus those of all the powers of the matrix. One might consider, therefore, the column sums of

$$T = A + A^2 + A^3 + \cdots + A^k + \cdots = (I - A)^{-1} - I \qquad [2]$$

where A is the original sociometric matrix with zeroes in the diagonal and I is the identity matrix.

Katz suggests, further, that indirect links be weighted inversely to the number of links involved. To do this a constant c is employed with $0 < c < 1$. The following formula then gives the matrix of summed and weighted values:

$$T = cA + c^2A^2 + \cdots + c^kA^k + \cdots = (I - cA)^{-1} - I \qquad [3]$$

Katz derives an equivalent formula for the computation of status that finds the solution through a set of linear equations rather than by finding the inverse of a matrix explicitly.

$$\left(\frac{1}{c} I - A'\right) t = s \qquad [4]$$

where A' is the transpose of A, and S is a column vector whose entries are the column sums of A. The formula yields a set of linear equations which may be solved for t, the sums of the columns of T. Katz also presents a formula for the weighting of t according to the number of possible choices in the group.

Leontieff Matrices

Hubbell, by working explicitly with the inverse matrix, develops

much more extensive information from an approach similar to Katz. Hubbell points out that matrices summarizing relationships within a group have been used by Leontieff in his input-output or interindustry models. He therefore transfers the Leontieff techniques to the sociomatrix.[3] The techniques are aimed at tracing the long-run effect of each member upon the others. In order to do this, use is made, as Katz does above, of the fact that $(I-A)^{-1}$ gives the sum of all the powers of A.

An example of the approach presented by Hubbell is the following: Suppose the matrix in Table 1 summarizes the relationships in a group. If the positive entries in the matrix are all changed to .25, this is the same as multiplying the matrix[4] by Katz' constant, $c=.25$. The inverse of $I-cA$ is given in Table 5. The inverse matrix can be taken to summarize the eventual effect of each member on any other. Thus, one unit of activity on the part of b will eventually give rise to .40 units by a. The row sums indicate the long-run influence of each member. In this case, the row sum for b is the largest and that for a is the smallest. The column sums, here equal, may be taken to signify the amount of constraint or pressure put on each member in the long run.

Up to this point, the work is the same as that outlined by Katz (1953) with one major difference. The complete matrix of long-run influence is obtained. (Other differences such as not subtracting the identity matrix from the inverse matrix are unimportant.) The use of the inverse matrix is significant for two reasons. First, it becomes easier to change the analysis from row sums to column sums. Second, it becomes possible to estimate the effect of various distributions of input from the external environment. Thus, if each member receives one unit of external input which he relays to members of his group, the eventual effect of each member can be computed by postmultiplying $(I-cA)^{-1}$ by a column vector of ones.[5] The column vector obtained is (1.51, 2.40, 2.09, 1.97) indicating that b is most influential. The effect of differences in the amount of external input can be similarly computed by postmultiplying with the appropriate column vector containing a different distribution of inputs.

It is also possible to assign differential weights to the likelihood of influence (or communication) travelling along a particular channel. For example, it may be that c pays more attention to a than to b. In this case, values in cells a_{13} and a_{23} might be .42 and .08, respectively, instead of .25. When the inverse of I minus this new matrix is computed, it is found that the effects of the change are to increase the long-run influence of a and d and decrease that of b and c.

TABLE 5

MATRIX OF $(I-cA)^{-1}$

1.05	.09	.28	.09
.40	1.20	.40	.40
.19	.38	1.14	.38
.36	.32	.17	1.12

[3] "Sociomatrix" will be used here and below instead of the longer "sociometric matrix."

[4] This weighting could be rationalized by assuming that each member of the group gives only half his attention to what goes on within the group and that he distributes his attention equally among the inputs from other group members. The column sums are all, therefore, taken to equal .50, and the values in the cells changed accordingly.

[5] This simply obtains the row sums discussed in the preceding paragraph.

This technique is therefore of special interest in analyzing groups with respect to the interrelation of the members and the possible effects of changes in demands, pressures, or environmental contacts. It may be possible to pick out cases in which the imposition upon group members of a given distribution of demands from the environment so overloads some members of the group as to cause a breakdown. Changes in the distribution of demands or the structure of the group may obviate this overloading.

Enumeration of Structures

Work has been done in counting the number of graphs or matrices that display specified characteristics such as a given set of row and column totals. The work on enumeration is important in obtaining chance distributions of various types of structures. The work cited in this section concerns only matrices with binary entries.

Number of Matrices with Given Row and Column Sums

Katz and Powell (1954) attack the following enumeration problem: Given a matrix with certain row and column totals, how many distinct matrices with the same set of row and column totals in the same order may be generated?

Solutions have been worked out and tables constructed by Sukhatme (1938), and by David and Kendall (1951), for matrices with binary entries. In the case of the sociomatrix, a restriction is usually introduced. The entries on the diagonal are either all zeroes or all ones. Katz has developed a formula that permits the use of the tables for these special cases.

The number of matrices (η) with zeroes in the diagonal cells, and a fixed set of column sums (s) and row sums (r) is given by

$$\eta(s, r) = A\left\{\prod_{i=1}^{n}(1+\delta_i)^{-1}(s, r)\right\} \quad [5]$$

The operator δ_i reduces entry i of the vectors of row and column sums by one. A is the number of all matrices, unrestricted with respect to the diagonal, that can generate the given set of column and row sums.

The formula for $\eta(s, r)$ is easy to apply since powers of δ have to be considered only up to m_i where m_i is the smaller of (s_i, r_i). This is expanded and the individual terms are evaluated using the tables mentioned above to evaluate the various A's obtained through the formula.

Number of Distinct Structures

A different enumeration problem has been attacked by Davis (1953). Given a set of n elements, how many distinct structures of relationship are possible between them? Two structures are distinct if they are not permutations of each other. In terms of matrices, two matrices are distinct if one cannot be obtained from the other by a simultaneous permutation of rows and columns. Davis develops formulae not only for the counting of the number of distinct structures, but also for the counting of specific kinds of relation structure, e.g., reflexive, symmetric, asymmetric, etc.[6] The bounds for the number of structures are given as follows:

$$2^{n^2-\log n/\log 2} < \text{number of distinct structures} < 2^{n^2} \quad [6]$$

Davis, in a later paper (1954), ap-

[6] Davis' paper concerns n-adic as well as dyadic structures. Only the dyadic structures are considered here. See also Copi and Harary (1953).

plied his formulae for the counting of structures to the special case of dominance relations.

Distribution of Subgroup Configurations

Closely related to the enumeration of group structures is work on the distribution of various subgroup structures. A large number of papers deal with the expected number and the distributions of various configurations on the basis of chance. This work has been done for stars (wheels), chains, rings (circles), various types of cliques and isolates. Much of this work has been carried forward by Katz (1952a; Katz & Olkin, 1952; Katz & Powell, 1957). It has importance in evaluating the results of the analysis of configurations. For example, if a given group has three isolates, it is of interest to discover how often this could have arisen by chance. However, a psychological theory or rationale to dictate the choice of configurations for study and to indicate why there should be departures from chance ordering has not yet been developed.

If and when theory and experimental investigation develop in which group structure is a dependent variable, then the distribution of particular configurations will have considerable importance. At the present time, however, configurations and their distribution seem to have little theoretical or practical significance.

Comparison of Groups

The question of the similarity of two matrices is important for either the evaluation of the amount of change in a group or the departure of the actual pattern from an ideal or required pattern. The techniques presented thus far for measuring similarity are basically correlational. They involve pairing of corresponding entries in two matrices and correlating the values found.

Comparison of Matrices Using Cell Entries

When a matrix has binary entries there are only four ordered pairs possible for the corresponding cells: (0, 0), (1, 0), (0, 1), and (1, 1). Katz and Powell (1953) therefore use a fourfold table to summarize the agreements and disagreements of entries in two binary matrices. They then construct an index of conformity based upon the observed cell frequencies:

$$\hat{\Gamma} = \frac{1}{n_A n_{\bar{B}}} [n(n-1)n_{AB} - n_A n_B] \quad [7]$$

where $n(n-1)$ = number of off-diagonal cells, n_A = number of positive entries in matrix A, n_B = number of positive entries in B, $n_{\bar{B}}$ = number of zero entries in B and n_{AB} = number of cells positive in both A and B. The index ranges between 1 and $-(n_B/n_{\bar{B}})$ and equals 0 when A and B are independent. The $\hat{\Gamma}$ for the agreement of A with B and B with A is not necessarily the same. For situations in which it is not assumed that one of the matrices is antecedent, the geometric mean of the two possible indices may be computed. The index is called C, a coefficient of concordance.

$$C = \sqrt{\hat{\Gamma}_1 \hat{\Gamma}_2} = \frac{n(n-1)n_{AB} - n_A n_B}{\sqrt{n_A n_{\bar{B}} n_B n_{\bar{A}}}} \quad [8]$$

It ranges between $+1$ and -1.

Katz and Powell point out that the same approach can be used to compare an individual's choices within matrices A and B (both describing, of course, the same group). A fourfold table is constructed on the basis of two corresponding rows of the matrices and the same indices computed.

Comparison of Matrices Using Row or Column Sums

Another type of comparison of matrices can be made by focusing on the row of column sums. This is easily accomplished by computing a product-moment correlation coefficient for the paired row sums or the paired column sums.

Hohn (1953) has discussed in detail the comparison of sociomatrices for the special case in which each individual ranks all others and zeroes are inserted in the diagonal. For this case the product-moment correlation between column sums may be written as follows:

$$\theta = \frac{\sum_i s_i s_i' - \frac{n^3(n-1)^2}{4}}{\sqrt{\left(\sum_i s_i^2 - \frac{n^3(n-1)^2}{4}\right)\left(\sum_i s_i'^2 - \frac{n^3(n-1)^2}{4}\right)}} \quad [9]$$

where s_i and s_i' are the column sums for the ith column of two matrices.

Similar correlations could be obtained for row sums or column sums in the case of any type of sociomatrix, i.e., matrices with any type of weighted entries. In these cases, however, the general formula for the product-moment correlation coefficient should be used.

Generalization of Correlational Approach

Katz (1947) has outlined a general method for evaluating agreement between the patterns of individual choices. The measure of agreement considered is based on the angle between vectors of choices; such as (0, 0, 1, 0) the first row vector in Table 1 (when zero is inserted in the diagonal cell).

The cosine of the angle between two vectors α and β is equal to their inner product, divided by the product of their magnitudes.

$$\cos \theta_{\alpha\beta} = \frac{\sum_i a_i b_i}{\sqrt{\sum_i a_i^2}\sqrt{\sum_i b_i^2}} \quad [10]$$

When a and b are deviations from their respective means, this is the standard correlation coefficient. To make use of either, it is necessary to assign values to the diagonal entries in the matrix. (It would probably be best to assign the highest possible positive cell value to all cells in the diagonal.) It is then possible to measure the agreement between the vector of responses generated by each individual in the group with that of every other member of the group.

Analysis of Subgroups

A large group often is assumed to have distinguishable subgroups within it. In certain organizations these subgroups are officially designated, e.g., the board of directors of a company. In other organizations, they are neither official nor immediately obvious to the outsider, e.g., factions within a political party. There have been several attempts to move from the intuitive grouping of individuals to a system for the unique segregation of subgroups on the basis of clear-cut criteria. The general interest of such a procedure is evident. In the analysis of man-machine systems, for example, it would be desirable to have a straightforward technique for the identification of natural

groupings (of men and machines). It is also of interest to the student of administrative organization, for example, to compare officially designated subgroups with the actually functioning subgroups in an organization.

Matrix Manipulation and Reduction Method

There are several techniques available for the analysis of subgroups. The first one suggested was that of Forsyth and Katz (1946) for matrices with bipolar entries ($a_{ij}=0$ or $+1$ or -1). It can also be applied to the simpler case with binary entries ($a_{ij}=0$ or $+1$).

The Forsyth-Katz technique involves simultaneously switching the rows and columns of the matrix so that positive entries cluster about the diagonal, i.e., their distance from the diagonal is minimized. If there are negative entries their distance from the diagonal is maximized. The technique helps identify cliques (clusters), leaders (center of cluster), and rank of cliques. The steps may be summarized as follows:

1. Place in Rows 1 and 2 (and Columns 1 and 2) a pair of individuals who choose each other.
2. Place in Row 3 (and Column 3) an individual who has mutual choices with both or, failing that, is chosen by both of the preceding pair.
3. Continue adding individuals in this fashion, using as criterion the requirement that each new member is chosen by at least 50% of the people already included in the subgroup.
4. Remove the subgroup when no more can be added and repeat the process with the reduced matrix.

After all the subgroups have been selected in this fashion, Forsyth and Katz specify methods for arranging the entire group along the diagonal (so that subgroups that reject each other are most widely separated) and for placing individuals who have not been included in any subgroup. The final manipulation is to arrange the members of each subgroup so that those receiving the greatest number of choices are at the center. The Forsyth-Katz technique requires rather awkward manipulation of matrices. Katz later developed a punched card technique for this analysis (1950). Methods for expediting the rearrangements can also be adapted from the scalogram board technique developed for Guttman's scalogram analysis.

Diagonal Maximization Method

Beum and Brundage (1950) have developed a more systematic technique for carrying out the rearrangement of a matrix with positive weighted entries. The technique is based on the following idea.

"It can be shown that if weights are assigned to the rows of a sociomatrix . . . and the average product of the elements in each column and the corresponding weights is maximized for each column, the sum of the squares of the elements about the principal diagonal is minimized." The technique as outlined below requires that the system of weighting individual choices or relationships gives the heaviest weight to the most important choices. The steps are as follows:

1. Zeros are inserted in the diagonal cells and column sums are obtained.
2. Weights of 1 to N are assigned to the rows of the matrix, beginning with 1 as the weight of the bottom row.
3. All matrix entries are multiplied by their row weight and weighted column sums are obtained.
4. The weighted column sums are divided by the unweighted column sums to obtain the average weight.
5. The matrix is then rearranged in order of average weights so that the column with the largest average is moved to the extreme left and the corresponding row is moved to the top. The column with

the next largest average is placed second from the left, etc.

The procedure is repeated until further iteration does not change the arrangement or results only in an alternating arrangement of the columns. The procedure is related to that used in obtaining a latent vector of the matrix. Its success depends on the absence of overlap in the subgroups.

Matrix Multiplication Method

A technique designed for the analysis of subgroups in binary matrices has been presented by Festinger (1949) and developed extensively by Luce and Perry (1949). The technique as first proposed analyzed "cliques." A clique was defined as a subgroup of three or more members all of whom were symmetrically related. The clique includes all members who meet the requirement of symmetrical relationship. For example, an individual is a clique member if he communicates with every member of the group and every member of the group communicates with him. This stringent requirement is relaxed in a later development by Luce (1950).

In this method, the matrix summarizing the group's relationships is reduced to an S-matrix consisting of all entries which are symmetric about the diagonal. This means eliminating all but mutual choices from the original matrix. The new S-matrix is then squared and cubed. The diagonal entries of S^2 give the number of mutual friends for each individual. The diagonal entries of S^3 indicate whether an individual is a clique member. The magnitude of the off-diagonal terms in both S^2 and S^3 indicates the compactness of the entire group.

The following theorems are presented and proved by Luce and Perry.

1. The magnitude of the cell entries $a_{ij}^{(n)}$ in A^n indicate the number of distinct n-chains from i to j.[7] An n-chain is a set of $(n+1)$ interrelated (interconnected) members. Thus, a speaks to b who speaks to c is a 2-chain.
2. A group member i has a main diagonal value of m in A^3 if, and only if, he has a symmetric relationship with m members.
3. A member i is contained in a clique if, and only if, his main diagonal entry in S^3 is positive.
4. If there are t diagonal entries in S^3 each of which equals $(t-2)\ (t-1)$ and the remaining diagonal entries are zero, then the t individuals form a clique.

Luce and Perry also present a general formula relating the magnitude of entries for member i in the diagonal of S^3 to the number, overlap, and size of the cliques of which i is a member. The formula and its proof are, however, incorrect. They apply only to special conditions of overlap in cliques.

On the basis of the theorems, the following procedure may be used to analyze a group:

1. Find the row in S^3 with the smallest or one of the smallest diagonal entries.
2. Choose those members whose entries for that row are the highest. The number chosen will depend on the size of the diagonal entry. (If the diagonal entry were $12=(5-2)\ (5-1)$ then the 4 highest entries would be selected.)
3. Examine the set selected in the S-matrix. If they all have positive entries in each other's row, then they form a clique.
4. Repeat the process with the next highest (or equal) diagonal entry until each of the clique members has been assigned to a clique.
5. Check the final set of cliques to make sure that each clique member is included in all cliques to which he belongs.

Luce and Perry also present theorems demonstrating that the properties of S are not lost in the process of cubing it, and a method for the computation of redundant 3-chains. A redundant n-chain is one in which the same individual appears more

[7] This relationship is the basis for Katz' status index. See Table 4.

than once. Thus, a redundant 2-chain would be *a* speaks to *b* who speaks to *a*. The work on redundancies, which is necessary for the identification and enumeration of chains and circles in the group, has been carried further by Katz (1952b) and Ross and Harary (1952).

Generalized Matrix Multiplication Method

In a subsequent article, Luce (1950) generalized the method by relaxing the definition of clique. He defined an *n*-clique as a subgroup of individuals who are all *n* or fewer links (choices, communication links) from each other. This extends greatly the utility of the approach.

The matrix manipulation for the analysis of *n*-cliques is closely related to the operations given above for the analysis of cliques. (They would now be called one-cliques.) Only the first few steps are different.

1. Compute the matrix powers A, A^2, A^3, $\cdots$ A^n.
2. Add these together.
3. Take the cells of the summed matrix, replace every positive entry with a one, and replace all diagonal entries with a zero.
4. From this point on, proceed as in the original matrix multiplication method.

The techniques developed by Luce have the advantage of requiring relatively simple procedures. They have one major disadvantage in that they severely limit the type of information they can handle. As developed, they can deal only with the presence or absence of a relationship, i.e., binary entries. Indications of degree of relationships (e.g., likes very much or communicates infrequently) cannot be handled. It is, of course, possible to reduce weighted entries to binary entries by dichotomizing them according to some criterion.

An example of the successful application of the matrix multiplication method may be found in a study by Chabot (1950). In the study, groups in an industrial situation were analyzed to test hypotheses concerning the relation of group membership to production.

Vector Analytic Methods

It was noted earlier that the inner product of vectors of choices can be used to compute a correlation coefficient. By taking each row vector that makes up the matrix of relationships, the matrix can be easily converted into a matrix of correlations. The entire body of factor analytic techniques can then be brought to bear upon the data.

Bock and Husain (1950) have presented another special type of factor analytic approach. They start with a choice matrix summarizing the subject's ranked preference for other group members. On the basis of the rank and whether or not the choices are mutual, weights are assigned to each cell entry. These weights are then analyzed by means of Holzinger's *B*-coefficient technique.

Rather than converting the vectors to correlation coefficients, and then following standard factor analytic procedures, it is more appropriate to deal directly with the original set of vectors that compose the sociomatrix. The steps then would involve setting up a basis for the space spanned by the vectors, and then rotating the basis to give some satisfactory fit.

Viewing the choice or communication vectors as an arrangement of vectors in a space leads to certain other ideas. In this arrangement, the number of dimensions in the vector space is related to the homogeneity of the group. A completely homogeneous group would have dimension equal to one. Everybody would choose everybody else, and would

have the choice vector (1, 1, 1, · · · , 1), if the diagonal entries are taken as equal to 1. All vectors would fall then on the equiangular line. The equiangular line gives then a baseline for measuring the homogeneity of the group. If the group is completely centripetal, then each individual chooses or communicates only to himself. The vectors are then in a n-dimensional Euclidean vector space with the basis (1, 0, 0, · · · , 0), (0, 1, 0, · · · , 0), etc. in which n = the number of individuals. It can be reasoned further that if the group consists of several subgroups, then this number will determine the dimensionality of the vector space. The number of dimensions required can be ascertained by solving for the number of nonzero latent roots of the matrix.

The advantages of a factor or vector analytic approach are the following:

1. The position of an individual in relation to all the subgroups or cliques may be described.
2. Use can be made of data in which different degrees of relationship are expressed.

It should be noted, however, that this type of approach involves a different definition of clique or subgroup. Individuals are members of the same clique if their patterns of relationship to other individuals are similar. It may be, therefore, that the Luce-Perry type of analysis is best suited for certain types of relationships (communicates with, influences) whereas a vector analytic approach is best suited for other relationships (chooses, finds himself similar to).

The practical implications of analysis of a group into subgroups should be underlined. The analysis is necessary whenever it is desired to group together individuals who interact (e.g., hand each other tools) most frequently. An example of this type of problem in an industrial setting may be found in the discussion by Chapanis, Garner, and Morgan (1949) on rearrangement of men and equipment in a shop. The application of the technique to other work situations, such as offices with communication problems, is obvious.

Assignment of Individuals to Subgroups

The analysis of subgroups is preliminary to the planned location or relocation of individuals in subgroups. It leads naturally, therefore, to the following general question. How can the individuals in a larger group be divided into subgroups on the basis of the relationships between the individuals?

Hotelling (1954) raised the issue in terms of the best way to assign individuals to a given number of teams, so that their sociometric choices determine their placement as much as possible. He set up the following procedure for computing the total satisfaction of the group under various groupings. First the matrix of choices is examined and then an assignment matrix J is constructed. The assignment matrix has entries 1 or 0 indicating whether an individual i has been assigned to team j. For example, given the choice matrix in Table 1, with zeroes on the diagonal, it may be desired to separate the group into two subgroups of two men each. Two possible arrangements are indicated in the assignment matrices J_1 and J_2 in Table 6. In J_1, a and b are assigned to one subgroup and c and d to the other. In J_2, a and d are placed together and b and c are placed together. The good fortune of individuals is defined in terms of $J'AJ$, the choice matrix postmultiplied by the assignment matrix and premulti-

TABLE 6

ASSIGNMENT MATRICES AND SATISFACTION MATRICES

	X	Y
a	1	0
b	1	0
c	0	1
d	0	1

J_1

	X	Y
a	1	0
b	0	1
c	0	1
d	1	0

J_2

$$J_1'AJ_1 = \begin{matrix} 1 & 3 \\ 3 & 1 \end{matrix} \qquad J_2'AJ_2 = \begin{matrix} 1 & 2 \\ 3 & 2 \end{matrix}$$

plied by the transpose of the assignment matrix.

The total satisfaction of the individuals in each group is indicated by the sum of the values in the principal diagonal. Thus J_2 with the entries on the diagonal summing to three would be considered a better assignment than J_1 where the $J'AJ$ diagonal sum is two. This means that in J_2 three out of the eight positive choices fall within the assigned subgroups.

Katz has begun work on situations in which the sizes of the subgroups are permitted to vary (1952b). Work has also begun on the development of techniques to maximize the sum of entries on the diagonal of the satisfaction matrix (Katz, Olkin, & Powell, 1952). This is a crucial problem because the number of ways to partition n individuals into k subgroups with m members each is $n!/(m!)k$. This easily becomes quite large. For example there are $12!/(4!)3$ or 34,650 ways to partition 12 individuals into 3 groups of 4 members each. This problem has similarities to the personnel classification problems that can be solved by linear programming.

OTHER APPROACHES: GRAPH THEORY LOGIC OF RELATIONS

The concern with methods of describing and analyzing groups has resulted in the attempt to bring a special branch of mathematics, graph theory, to bear on the problems of group organization. A nonmathematical presentation of graph theory was made in a monograph by Harary and Norman (1953). Most of the monograph is devoted to the discussion of definitions that have been employed in mathematical graph theory and the relation of the definitions to psychological concepts. For example, Lewin's concept of boundary between two regions in a life space is translated here as in Bavelas (1948, 1950) into a connection between two points.

In addition, Harary and Norman consider a generalization of graph theory to adapt it to psychological problems. The generalization includes the addition of such concepts as strength—the number of lines joining a pair of points. Strength, of course, has been discussed previously as weighted cell entries. Moreover, since the graph as defined in mathematical work centers on symmetric relationships, they also present the concept of the directed graph in which symmetry is not required.

Although some studies have been done using the vocabulary of graph theory, its effect on psychological theory and experimentation has not been extensive to this time. Examples of the use of the vocabulary may be found in the recent article by French on social power (1956) and in Weiss' work on organizational structure (1956). Out of the work in this area, however, other techniques have been developed for the analysis of subgroups.

A point that has been stressed on the basis of graph theory considerations concerns the role of key or liaison positions in a group. These are positions that serve as links between subgroups, i.e., positions whose elimination results in the group's falling into distinct sub-

groups. Ross and Harary (1955) have developed techniques for determining the liaison positions in a group. The steps in the analysis are the following:

1. A symmetric matrix with binary entries is written describing the group's structure. Note that here again there is the restriction to symmetric matrices.
2. A matrix of distances between positions is constructed. This can be computed by using powers of the original matrix. Zeros are entered in the diagonal and one is entered in each cell with a positive entry in the original structure matrix A. Successive powers of A are then taken. Whenever a cell entry first increases from zero, then the power of the matrix in which this change occurs is entered in the corresponding cell of the distance matrix. (See Luce and Perry's first theorem above.) For example, if a_{ij} becomes positive when the matrix is squared, a two is entered in the corresponding cell of the distance matrix. Successive powers of the matrix are taken until all cells are accounted for.
3. The highest number in the distance matrix is found and the positions whose columns contain this maximum number are eliminated from consideration as liaison positions. (These are called peripheral points.)
4. Positions whose entry in any row is the highest for that row are eliminated. (These are called relatively peripheral points.)
5. Positions are liaison positions whose entries in any row satisfy the following requirements:
   (*a*) not zero,
   (*b*) not the maximum for that row,
   (*c*) unique for that row.

These steps may still leave some positions unclassified. Ross and Harary present techniques for classifying the remaining positions. The discovery of liaison positions can have two uses. One is to distinguish key personnel in an organization. The other is to afford a basis for analysis of the group into subgroups or cliques. Harary and Ross (1957) have also presented another method for the analysis of cliques in binary matrices that covers the case of overlapping subgroups. In order to do this they introduce the operation of elementwise multiplication of matrices symbolized by X. Under this operation, each cell in one matrix is multiplied by the corresponding cell in the other matrix. The analysis is based on the use of individuals who are members of only one clique, "unicliqual" individuals. The steps in the analysis are the following:

1. The symmetric matrix, S, is obtained by eliminating nonreciprocated cell entries.
2. The matrix S^2XS is constructed. The cell entries indicate whether a pair of individuals belong to the same clique.
3. All rows which consist entirely of zeros are eliminated. Their corresponding columns are eliminated. The reduced matrix is called M.
4. The row sums, r, and the number of cell entries, n, greater than zero in each row are obtained for M. If, for a given row
$$r = n(n-1)$$
then the individual represented by that row is member of only one clique. All members of that clique can be isolated by selecting all other individuals with a positive entry in that row.
5. If this clique does not include all members in M, this procedure is carried further. The group is divided into two subgroups: those who belong only to the first clique, and the remainder. The procedure described above is repeated on the remainder subgroup.

After all unicliqual members are accounted for, further subgrouping and analysis is continued. These subgroups are defined on the basis of members whose row sums are minimal in the last reduced matrix.

In addition to the techniques and approaches discussed thus far, recent developments in the fields of mathematics and logic promise to be relevant. It may be that the study of groups requires consideration of more complicated relationships than the dyadic relationships considered above (e.g., a talks to b about what c said to d, or as in some of Tagiuri's recent work (1952)—a expresses his attitude to b, and predicts b's attitude toward himself). Copilowish (1948),

Copi and Harary (1953), and Davis (1953) have considered such *n*-adic relationships.

Discussion

One of the major outcomes of the work described above has been to present a set of techniques for simplifying and analyzing the complex data generated by group functioning. These techniques, moreover, involve the translation of the data into mathematical form that permits the application of a wide range of powerful analytic techniques. In addition, these techniques are applicable not only to the original sociometric relations (e.g., "like," "chooses") but also to other relations (e.g., "communicates to," "hands materials to," etc.) They therefore permit the analysis of a much wider variety of organizational relationships than hitherto studied.[8]

As these analytic techniques become better known and more widely applied they will help promote the construction of quantitative systems for group behavior. It should not be expected, however, that these or related techniques will furnish easy solutions to the problems of group structure and functioning. In most cases they will probably be helpful only in clarifying the requirements for adequate descriptive or explanatory systems. At best, these techniques will provide the variables to be incorporated in such theories. There remains the task of constructing theories concerning the behavior of groups. There are two types of theorizing possible with sociometric or structural data. One concerns the causes of particular patterns or indices. The other concerns the effects of these. An approach to theories concerning the cause of patterns is indicated in some work by Landau (1951a, 1951b, 1953) on the special case of dominance relationships. Much more work of this type is needed before the usefulness of the various measures suggested in the sociometric literature can be assessed. The second type of theorizing concerns the effects of patterns. This type of theory initiated the group network studies. These studies will be evaluated in a subsequent paper.

[8] Examples of the structural properties of formal military teams that may be submitted to this type of analysis are considered by Glaser (1958).

REFERENCES

Bavelas, A. A mathematical model for group structures. *Appl. Anthrop.*, 1948, **7**, 16–30.

Bavelas, A. Communication patterns in task-oriented groups. *J. acoust. Soc. Amer.*, 1950, **22**, 725–730.

Beum, C. O., & Brundage, E. G. A method for analyzing the sociomatrix. *Sociometry*, 1950, **13**, 141–145.

Bock, R. D., & Husain, S. Z. An adaptation of Holzinger's B-coefficients for the analysis of sociometric data. *Sociometry*, 1950, **13**, 146–153.

Chabot, J. A simplified example of the use of matrix multiplication for the analysis of sociometric data. *Sociometry*, 1950, **13**, 131–140.

Chapanis, A., Garner, W. R., & Morgan, C. T. *Applied experimental psychology—human factors in engineering design.* New York: Wiley, 1949.

Copi, I. M., & Harary, F. Some properties of *n*-adic relations, *Portug. math.*, 1953, **12**, 143–152.

Copilowish, I. M. Matrix development of the calculus of relations. *J. symbol. Logic*, 1948, **13**, 193–203.

Criswell, Joan H. Notes on the constant frame of reference problem. *Sociometry*, 1950, **13**, 93–107.

David, F. N., & Kendall, M. G. Tables of symmetric functions, Parts II and III, *Biometrika*, 1951, **38**, 435–462.

Davis, R. L. The number of structures of finite relations. *Proc. Amer. math. Soc.*, 1953, **4**, 486–495.

Davis, R. L. Structures of dominance relations. *Bull. of math. Biophysics*, 1954, **16**, 131–140.

Edwards, Daisy S. The constant frame of reference problem in sociometry. *Sociometry*, 1948, **11**, 372–379.

Festinger, L. The analysis of sociograms us-

ing matrix algebra. *Hum. Relat.*, 1949, **2**, 153–158.

FORSYTH, E., & KATZ, L. A matrix approach to the analysis of sociometric data: Preliminary report. *Sociometry*, 1946, **9**, 340–347.

FRENCH, J. R. P. A formal theory of social power. *Psychol. Rev.*, 1956, **63**, 181–194.

GLASER, R. *Current Trends in the Description and Analysis of Behavior*, "Descriptive variables for the Study of Task-Oriented Groups," Univer. of Pittsburgh Press, 1958.

HARARY, F., & NORMAN, R. Z. *Graph theory as a mathematical model in social science.* Ann Arbor, Univer. of Michigan, 1953.

HARARY, F., & ROSS, I. C. A procedure for clique detection using the group matrix. *Sociometry*, 1957, **20**, 205–215.

HOHN, F. E. *Some methods of comparing sociometric matrices.* Bur. of Res. and Serv. Tech. Rep. No. 5, Urbana: Univer. of Illinois, 1953.

HOTELLING, H. Research in multivariate statistical analysis at Chapel Hill. Office of Nav. Res. Rep., Inst. of Statist., Univer. of North Carolina, 1948.

HUBBELL, C. H. Sociometry and Leontieff matrices. Univer. of Michigan seminar in applic. of math. Memo. 19, 1953.

JENNINGS, H. H. *Leadership and Isolation* (2nd ed.) New York: Longmans, Green, 1950.

KATZ, L. On the matric analysis of sociometric data. *Sociometry*, 1947, **10**, 233–241.

KATZ, L. Punched card technique for the analysis of multiple level sociometric data. *Sociometry*, 1950, **13**, 108–122.

KATZ, L. The distribution of the number of isolates in a social group. *Ann. math. Statist.*, 1952, **23**, 271–276. (a)

KATZ, L. Recent advances and some unsolved problems of sociometric theory. Statist. Res. Lab. Memo. 8, East Lansing: Michigan State Univer., 1952. (b)

KATZ, L. A new status index derived from sociometric analysis. *Psychometrika*, 1953, **18**, 39–43.

KATZ, L. A probability model for one-dimensional group organization. Univer. of Michigan seminar in applic. of math., Memo. 23, 1954.

KATZ, L., & OLKIN, I. Expectations of certain configurations in networks. Statist. Res. Lab. Memo. 7, East Lansing: Michigan State Univer., 1952.

KATZ, L., OLKIN, I., & POWELL, J. H. The relation of "satisfaction" to reduction of the choice matrix. Statist. Res. Lab. Memo. 2, East Lansing: Michigan State Univer., 1952.

KATZ, L., & POWELL, J. H. A proposed index of the conformity of one sociometric measurement to another. *Psychometrika*, 1953, **18**, 249–256.

KATZ, L., & POWELL, J. H. The number of locally restricted directed graphs. *Proc. Amer. math. Soc.*, 1954, **5**, 621–626.

KATZ, L., & POWELL, J. H. Probability distributions of random variables associated with a structure of the sample space of sociometric investigations. *Ann. math. Statist.*, 1957, **28**, 442–448.

LANDAU, H. G. On dominance relations and the structure of animal societies: I. Effect of inherent characteristics. *Bull. of Math. Biophysics*, 1951a, **13**, 1–19.

LANDAU, H. G. On dominance relations and the structure of animal societies: II. Some effects of possible social factors. *Bull. of Math. Biophysics*, 1951b, **13**, 245–262.

LANDAU, H. G. On dominance relations and the structure of animal societies: III. The condition for a score structure. *Bull. of Math. Biophysics*, 1953, **15**, 143–148.

LINDZEY, G., & BORGATTA, E. F. Sociometric measurement. In G. Lindzey (Ed.) *Handbook of social psychology.* Vol. I. Cambridge, Massachusetts: Addison-Wesley, 1954. Pp. 405–448.

LUCE, R. D. Connectivity and generalized cliques in sociometric group structure. *Psychometrika*, 1950, **15**, 169–190.

LUCE, R. D., & PERRY, A. D. A method of matrix analysis of group structure. *Psychometrika*, 1949, **14**, 95–116.

MORENO, J. L. *Who shall survive? A new approach to the problem of human interrelations.* New York: Beacon, 1934.

PROCTOR, C. H., & LOOMIS, C. P. Analysis of sociometric data. In Marie Jahoda, M. Deutach, & S. W. Cook (Eds.), *Research methods in social relations: With special reference to prejudice.* Part Two. New York: Dryden, 1951. Pp. 561–585.

ROSS, I. C., & HARARY, F. On the determination of redundancies in sociometric chains. *Psychometrika*, 1952, **17**, 195–208.

ROSS, I. C., & HARARY, F. Identification of the liaison persons of an organization using the structure matrix. *Mgmt. Sci.*, 1955, **1**, 251–258.

SUKHATME, P. V. On bipartitional functions. *Phil. Trans.* **237A**, 1938, 375–409.

TAGIURI, R. Relational analysis: An extension of sociometric method with emphasis upon social perception. *Sociometry*, 1952, **15**, 91–104.

WEISS, R. S. *Processes of Organization.* Ann Arbor: Univer. of Michigan, 1956.

Received November 17, 1958.

PSYCHOLOGICAL BULLETIN
VOL. 56, NO. 5, 1959

MULTIPLE METHODS OF PERSONALITY ASSESSMENT[1]

RONALD TAFT
University of Western Australia

The term "personality assessment" refers to any procedure aimed at describing a person's characteristic behavior by categorizing him with respect to some communicable dimension or dimensions.[2] Since the OSS assessment procedures, however, the term has tended to be pre-empted for the procedure where several different types of assessment techniques are applied to the subjects and the final assessments are made by the combined judgments of several assessors concerning the subjects' predicted behavior outside of the assessment situation. These procedures are "multiple" in two senses: with respect to the techniques and with respect to the assessors.

Our treatment in this paper will deal with the basic logic of this type of assessment, and the discussion will be illustrated by the best known multiple personality assessments, details of which are outlined in Table 1.[3] Each one of these assessments has, in its own way, constituted a milestone in the history of multiple personality assessment.

The researches into personality conducted at Harvard in the 30's under the direction of Murray (1938) were the first to use the typical procedures of personality assessment—diagnostic committee assessments of personality based on interviews and a varied battery of objective, projective, and situational tests. However, unlike the later assessments, no outside criterion was used in these Harvard studies, and, therefore, no more than passing reference will be made to them. The same applies to the continuing series of studies of personality carried out by Cattell and his students (Cattell, 1957) which started to employ external criteria only at an advanced stage of its progress. The British War Officer Selection Boards (WOSB), which were inspired by the German officer multiple technique selection procedures (Farago & Gitler, 1941), pioneered the use of a quasi-natural social situation, including the leaderless discussion, as a basis for judging the potential social skills of the candidate. They also produced the first validation material on multiple assessment procedures as a means of selection. The British Civil Service Selection Boards (CISSB) continued this work, with more emphasis on the validation of individual techniques as well as the technique as a whole. The OSS assessment highlighted the psychological problems inherent in assessment and won many supporters for the value of combining multiple

[1] The author expresses his thanks to the colleagues who have discussed various points in this paper with him, especially to James Lumsden; also to Saul B. Sells for his valuable comments.

[2] This is the same procedure as "instantiating a person object in a module or set of modules," a terminology which the writer has preferred in another context (Sarbin, Taft, and Bailey, in press), but which is avoided here in the interests of communicability.

[3] Insofar as the assessments use multiple techniques, the problems of inferring the predictions and validating the tests are the same as those involved in other multi-variate procedures. See, for example, the treatment of these problems in Thorndike (1949). Our emphasis here will be mainly on the problems that arise from the combination of multivariate procedures and multiple assessors.

TABLE 1

DETAILS OF MILESTONE ASSESSMENT PROGRAMS

(Wide sample of techniques used: individual and multiple interviews; observation of group activities and situational tests; objective, projective, and performance tests; "made to measure" inventories)

Assessment	Date	Assessees	Primary purpose	Strategies (in order of importance)	Assessors
Harvard (Murray, 1938)	1934–37	Young men, mainly Harvard undergraduates (paid subjects)	Personality research	Analytic	Psychologists
WOSB (Harris, 1949; Morris, 1949)	1942–45	British officer candidates	Selection	Analytic Global	Army officers, psychiatrists, and psychologists
OSS	1944–45	U. S. Intelligence and espionage agent candidates	Selection	Analytic Global	Psychologists, psychiatrists, and other social scientists
Michigan, VA (Kelly & Fiske, 1951)	1946–49	Clinical psychology graduate students	Validation of techniques	Empirical Global	Psychologists (clinical and nonclinical)
California IPAR (Various published and unpublished reports, e.g., Barron, 1954; Gough, 1953)	1950–51	Advanced graduate students	Personality research; validation of techniques	Empirical Analytic Global	Psychologists
Chicago (Stern, Stein, & Bloom, 1956)	1952–54	Students in theology, education, and arts	Validation of techniques	Analytic Global	Psychologists
Menninger (Holt & Luborsky, 1958)	1946–52	Psychiatric training candidates	Selection, validation of techniques	Global Analytic Empirical	Psychiatrists and psychologists

tests and observations by pooling the judgments of several assessors; the Michigan VA assessment program did much to upset that support while the Chicago and Menninger assessments reinstated some of it through their promising findings. The California Institute of Personality Assessment and Research (IPAR) differs from the other assessments in emphasizing research into personality to a greater extent.

THE ORIENTATION AND PURPOSE OF PERSONALITY ASSESSMENTS

Three foci of assessment can be distinguished: human performance in some socially defined situation or situations (the criterion performance); performance in defined assessment situations, i.e. tests (the assessment performance); and the link between these two performances (test validation). Different assessment programs have been oriented towards one or more of these aspects depending on their primary purpose (see Table 1). The orientation towards criterion performance implies the primary purpose of assessing candidates with respect to the criterion in order to select or reject them. The orientation towards the assessment performance is concerned with the validation of the assessment techniques themselves, while the orientation towards the link between performances is concerned with research on the functioning of personality.

Selection was the original purpose for the WOSB, CISSB, and OSS assessments; in each case, the assessors were presented with the immediate problem of selecting from a given group of candidates those who would make the most adequate army,

TABLE 1 (*Continued*)

Criterion analysis method	Method of rating	Main criterion	Some selected validities (uncorrected for selection of groups)
—	Committee	No external criteria	—
Personal knowledge	1. Committee and 2. Final Review Board	Supervisors' reports	1. CISSB committee 0.13–0.25. 2. Review Board 0.23–0.41. (When corrected for selection the range of validities is 0.50–0.66.)
Intuitive and interview with experts	Committee	Field reports by the assessors and by field commanders on several molar traits	"Over-all" ratings 0.08–0.53 (varying with assessment group and criterion). Rating of "Effective Intelligence," 0.33–0.53
Personal knowledge	Individual and pooled ratings	Ratings by clinical teachers and supervisors on several aspects of clinical work	"Over-all" rating and clinical competence, 0.37. Miller Analogies and clinical competence, 0.35. Strong Interest Key for Clinical Psychological and Research Competence, 0.35. Other validities lower
Personal knowledge	Committee	1. Teacher's prediction of student's professional potential. 2. Teacher's ratings of personal soundness	1. Cross-validated inventory with 1., 0.29 2. Committee ratings with 2., 0.41
Committee job analysis and interviews with teachers	Committee	Teacher's judgments and exam results	Very high validities
Committee job analysis and success and failure	Individual and averaged ratings of interviewers	Supervisor's ratings (pooled) on specific and general competence	Interviews (global), 0.24. Interviews (analytic), 0.26. Tester's analytic ratings on projectives, 0.27. Objective scoring of projectives cross-validated at zero. Best interviewer (all data) 0.57

civil service, and secret service officers, respectively. After a consideration of the personality requirements of the positions for which they were selecting officers, the assessors judging the candidates on the basis of techniques chosen either because they appeared to have face validity for measuring these requirements or because the assessors were familiar with their use. At least in the case of wartime assessments, neither the time available nor the conditions permitted more scientific procedures than that, and it was hoped that accuracy would be achieved through weight of numbers (of techniques and of assessors).

Test validation (and construction). Some of the later assessments, notably the VA study, set as their short-run aim the task of developing and validating techniques for future use in selection. The validation studies were applied not only to the individual items and tests, but also to the purely subjective techniques, such as group observations and interviews. In some of them, e.g., Menninger, the individual judges were also validated as though their judgments were scores on a test. When we speak of the validity of assessment techniques it is important to include these judgments among the techniques, as they vary greatly in their accuracy.

The Harvard studies were the first to use multiple assessment techniques for *personality research*, and the outstanding recent example is the IPAR work at California. (The large-scale factor analytic studies of personality, e.g., Cattell and Eysenck [1947], did not use the combined judgments of several assessors.) The

Harvard studies dealt with the correlations between different performances that were elicited in the test situation, whereas the IPAR studies were concerned, in addition, with the relationship between the assessment performances and criterion measures such as ratings by university teachers of the subject's professional potential, his originality, and his personal soundness. The Michigan studies of clinical psychologists were similar in orientation.

Most of the personality assessments have tried to pursue more than one of the above purposes at once, but there are drawbacks to such attempts at economy. For example, an attempt was made in the CISSB studies (Vernon, 1950) to combine selection and validation, but the validation indices were lowered and distorted by the attenuation of the sample through rejection of candidates. The low validities obtained became remarkably high (for that sort of prediction) when a correction for selection was applied, but such corrections are only arbitrary estimates. The use of assessment procedures for selection implies that the procedures have already been validated, but this has usually not been the case. The assessors have either had to use whatever prior knowledge they possessed about the validity of the techniques for the purpose at hand, or they have had to base their predictions on the relevant postulates in their theory concerning the link between the assessment and the criterion behavior of the subjects. For example, the assessors presume that the situational tests in the assessment program have what Cronbach and Meehl have termed "content" validity (1955). But in selection, this type of validity can be regarded only as a holding procedure for an ultimate "predictive" validity. Where the criteria are imprecise and not repeatable, or where selection is urgent, a separate validation study may not be practical, and under these circumstances there is no alternative to conducting selection without prior validation. It still may be possible, however, over a period of time, to utilize the imperfect validational material that becomes available in order to improve the existing selection procedures. This seems to have been the case, for example, in the OSS studies.

Validation studies of the assessment techniques also logically precede the use of those procedures for personality research, although techniques used in such research often are accepted on the basis of their face validity. To use the one and same study to validate the techniques and to use them to measure personality is lifting oneself up by one's boot-straps. In fact, however, the assessments which attempt to carry out this dual purpose obtain independent support for the "bootstrap lift" from already existing information regarding both validation and the functioning of personality. Even then, the interpretation of personality research projects that do not commence with a pilot study on the validity of the instruments is always subject to doubt. How do you know that expressed hostility to authority figures on the TAT measures suppressed rebellious tendencies? How do you know, when assessor X observes a subject to be dominant, that he is dominant? How do you know that observed "role empathy" in a role-playing test is a valid predictor of social skill? Such questions can be answered only by the progressive refinement of validity information and personality theory.

Assessment procedures usually rely on many unvalidated tests, and when the correlations between the tests are

used as a means of studying personality—as in the case, for example, of the Harvard and IPAR studies—it is necessary to decide whether these correlations are to be treated simply as validity indices, or whether the validity of the tests will be assumed and the correlations treated as throwing light on the relationships between different personality structures. The problem of simultaneous validation of tests and the study of personality is related to the problem of "concurrent" and "construct" validity. By setting up some of the behavioral measures made during the assessment as tentative criteria, it is possible to validate other assessment measures against these. Cronbach and Meehl call this concurrent validation, and it is one way of utilizing previous knowledge of validities by choosing criteria measures that have reasonably well-established reliabilities and validities. Then on the basis of all that is known about these measures, their implications for the understanding of personality can be explored further by a strategy of construct validation. The data collected during the assessment can be added to the "nomological nets" already used in thinking about the particular personality constructs and new hypotheses developed for investigation in later studies. Thus, even an assessment program that is aimed primarily at the purpose of selection can make a contribution to personality research through construct validation. (The place of construct validity in an assessment program is discussed more specifically below under the heading of *analytic strategy*.) This concept also enables an assessment program to avoid the problem of the priority of validation of instruments (versus conducting personality research) by conceiving both validation and personality research as two aspects of the one endeavor, both aspects gradually throwing light on each other as more and more data accumulate.

But this double-aspect approach of construct validation is an uneconomical process. Refinements may often be made more readily to our personality theory or to our knowledge of the validity of the techniques by a more direct approach to one or the other. In this case the problem of priorities which we have discussed cannot be avoided.

The Prediction Strategies in Assessment

The Criterion

All assessment programs involve studies of the link between two or more pieces of behavior, whether the primary purpose be selection, validation research on tests, or personality research. Some of this behavior is known as assessment behavior and some as criterion behavior. These concepts are analogous to the independent and dependent variables in experimental psychology, and it is an arbitrary decision by the experimenter which one is designated as which. Most of the reports of assessments have devoted some space to the criterion problem, especially the report of the Chicago assessments (Stern, Stein, & Bloom, 1956). Most of the problems are similar to those involved in the validation of multivariate objective techniques discussed, for example, by Thorndike (1949).

A special problem that arises in personality assessment is the frequent unreliability of the criteria which so often represent subjective judgments that vary from one criterion rater to another. This unreliability imposes a serious limitation on the potential validity of personality assessments, and it makes it

difficult to evaluate some of the low validity coefficients reported.

The designation of the criteria of performance is determined by the circumstances of the assessment, and usually must be taken for granted by the assessors. Thus, in the Chicago study the assessors explicitly accepted the principle that the criterion ratings represented the predilections of one or more supervisors with whom the subjects interacted in the criterion situation, and that the assessors' predictions of the subjects' success must be made in reference to the "psychological job requirements" implied by these predilections and interactions. The assessment strategy should be aimed at the criterion, once the latter has been established. Kelly (1957) did not accept this principle in his researches on medical school selection. In this study he analyzed the criterion measures and found that there were at least three, and possibly four, types of medical performance which could be predicted independently. In the long run, however, a selection program has to choose between the independent criteria, or the criteria have to be combined by some type of simple, weighted, or complex, interactional summation, or by taking account of one critical instance.

A complication that arises in criteria analysis, such as that of Kelly, is that an assessor can only predict to indices of the criteria, not to the actual criteria themselves. It may be possible in some instances for the assessor to demonstrate that an index used in assessment has a low correlation with some more satisfactory, although less accessible, criterion index; for example, that academic grades in medicine do not represent the doctor's subsequent service to the community as a practitioner. Assuming that the latter is accepted as the more fundamental in medical practice, the assessors should predict to it rather than to academic grades by trying to obtain some accessible index which more realistically measures this criterion of community service. Sometimes the assessors may be able to convince those who control the criterion ratings that the indices which the latter are using are not consistent with their fundamental criterion, but eventually the assessors and the criterion raters must agree on some criterion index in accordance with the policy of the organization. Otherwise it would be absurd to speak of the validity of the assessment.

Three types of strategies can be distinguished for predicting the criteria performance: *naive empirical*, *global*, and *analytic*, and we shall now consider each strategy in detail.

1. *Naive empirical.* This refers to the classical method of test construction, adapted from aptitude testing, in which the inclusion in a selection program of a test—or test item, which may be treated for our purposes as a separate test—is determined mainly by its predictive validity, i.e., by the degree to which it correlates with or discriminates a specified criterion. Tests that are not sufficiently valid are either dropped from the program or amended and no consideration is given to the meaning of the test behavior, except as an afterthought. The naive empirical strategy, thus, is one in which inference proceeds directly from test to criterion without the mediation of intervening variables.

Not a great deal of use has been made of this empirical strategy in multiple personality assessment, partly because of intellectual resistance to atheoretical procedures on the part of personality researchers, and partly because of the absence of reliable criteria. The outstanding ex-

amples of the use of the naive empirical strategy are found in the IPAR studies, especially in the scales of the California Psychological Inventory and the Adjective Check List compiled by Gough. These scales, which give unit weight rather than beta weights to the tests, i.e., items, enable predictions to be made to quite complex behavioral criteria, for example, tolerance, delinquency, academic achievement, neurodermatitis, potential social status (Gough, unpublished bibliography, IPAR, 1955).

The naive empirical strategy has the advantage over other strategies of objectivity and also of enabling assessors to predict complex and little understood behavior. But it also has serious limitations: it can be used only where suitable criteria groups are available for validation and cross-validation, and the validities may "drift" owing to changes in significant aspects of the conditions—temporal, geographic, public attitudes and information, set of the subjects, etc. Either some understanding of the underlying theoretical factors is necessary to provide a warning system against "drift," or constant revalidation must be carried out.

The primary purpose served by the naive empirical approach is that of constructing, and validating assessment instruments, although the long-range purpose can be both selection and research on personality. Up to a point, the personality research aim can be served simultaneously with the validation aim, since the discovery of the intercorrelations between the tests themselves and the criteria can suggest personality constructs. But we are now back on the problem of priorities: we can use validation studies for personality research only if we already possess postulates about the significance for personality of the behavior tapped by the tests and the criteria.

In this reference we should briefly consider the sources of the test items that are used in the validation "tryouts." The sources may be naive empirical, or they may be theoretical. Empirical sources include: tests in the general area that are traditionally used, those that are readily available and can conveniently be given, tests whose title or item content bear a superficial relationship to the criterion, and tests which have previously been shown to relate to the criterion. Theoretical sources of tests, on the other hand, include the systematic or unsystematic sampling —usually the latter—of the areas of personality that are considered by the researcher to be relevant to the criterion behavior. The empirical outlook of the student who is developing personality assessment techniques is seldom so naive that it is entirely uninformed by theoretical considerations, so that the "naive empirical" approach in practice tends to become mediated by intervening structures and thus to approach the analytic strategy described below. The intervening structures, however, are not made explicit in this empirical approach.

2. *Global.* This is the second nonmediated strategy, in which the assessor relies on his intuition, empathy, and *verständnis* processes to provide the predictions, rather than using statistically established associations between assessment behavior and criteria. If any analysis is made of the criterion in a global strategy it is directed at the social role expectations for the criterion performance rather than at the required personal qualities for successful performance. (The latter is more appropriate to the analytic strategy discussed below.) Information may be given to the as-

sessors about the subject's performance on objective tests, and even concerning the validity of these tests—for example in the Menninger studies (Holt, 1958)—but the ultimate assessment is a global one. This procedure is the personality assessor's answer to some of the drawbacks of the empirical strategy. Intuitive predictions can be used when the assessors have only a vague concept of the criterion conditions, but empirical methods require clear-cut criteria and expendable samples of trial subjects who have been rated on these criteria. Where this is impossible, as in the case of the OSS studies, the empirical strategy cannot be used, and intuition must be resorted to.

The distinction between empirical and global strategies also is analogous to the distinction between narrow- and wide-band techniques (Cronbach & Gleser, 1957), the former enabling comparatively more reliable but limited predictions. Supporters of the global strategy have claimed for it special adaptability to the vagaries of the conditions associated with both the assessment and criterion situations. Some writers also claim for it a special virtue in connection with personality research in that it avoids the violation of a "whole" person inherent in trait psychology; however, it is very doubtful whether it is correct to use the word "research" to describe a mode of study which, if it were applied in its pure form of global *verständnis*, would by definition preclude communication of the assessments.

The value of the claims of the global strategists to have improved on empirical validation as a basis for selection programs is limited. Subjective methods of making predictions have seldom been shown to be superior to objective methods where these are available, excepting in the case of especially competent assessors (see below). The relative competence of the assessors in making predictions about the subjects is analogous to the relative validity of the tests, and both can be established by the same type of validation techniques. In this way the empirical and the global strategies are similar in orientation: the proviso that incompetent assessors should either be eliminated from the assessment panel or trained to eliminate errors is analogous to the dropping or amending of an invalid test in the empirical strategy.

The "nonanalytic" techniques used in the global strategy are not necessarily nonmediated by personality constructs, even though these constructs may not be made explicit. The process of moving from observations of behavior to inferences about future behavior uses a set of postulates about personality and various derived premises; these premises involve certain personality constructs or categories into which the assessor places the behavior of the subjects, i.e., he "instantiates" the behavior. Intuitive inferences, even empathic ones, can be reduced to this formulation which provides a bridge between analytic and nonanalytic processes. This point is elaborated by Sarbin, Taft, and Bailey (in press).

3. *Analytic.* The analytic strategy makes explicit the role of mediating constructs in prediction. A two-stage inference is involved; first, there is an inference from the criterion requirement to the traits that are relevant to that performance (the "criterion analysis"); and, secondly, an inference from the subject's observed behavior and test performance to his status on the trait dimensions (the assessment). Research on the

validity of these inferences requires two separate studies: one of the validity of the analysis of the criterion requirements and the criterion indices, and one of the validity of the tests as predictors of the criterion. These validation studies should be based on independent samples of behavior and, for preference, on independent samples of subjects, the research on the criterion analysis to precede the validation of the instruments.

The importance of criterion analysis was recognized in each one of the "milestone" assessments, but the validity of the analysis is usually assumed. Two types of approach to the criterion analysis problem have been used: intuitive and empirical. The *intuitive* approach is the one usually used in personality assessment; typically the assessors have used either the testimony of "experts" or their own theoretical analysis to determine the criterion requirements. These analyses rest on a theory of personality, but the theory is usually not make explicit, nor is it subjected to empirical validation.

The empirical approach to criterion analysis can employ subjective or objective methods. The Menninger studies, for example, employed subjective rating methods to compare the characteristics of successful and unsuccessful psychiatrists.[4] The VA assessment program was, among other things, one big empirical criterion analysis using both subjective and objective methods. The study began with no explicit analysis of the requirements in clinical psychology and ended with an explicit description of some of the characteristics which relate to success in various aspects of that profession. In a sense, all preliminary validation try-outs of tests in a naive empirical strategy, such as those used in the VA and IPAR programs, constitute a criterion analysis. The cross-validation that follows may thus be regarded as testing a series of hypotheses about the criterion behavior. Referring once more to Cronbach and Meehl's contribution (1955), we see now that the analytic strategy is a type of construct validation which attempts to augment the "nomological net" surrounding the relevant constructs.

The main difficulty with the analytic method of assessment is that it requires a set of constructs which may not exist in our present state of psychological knowledge—although the assessment results may contribute to the development of such constructs. The difficulties which factor analysts often encounter in their attempts to label their factors leads one to sympathize with Cattell's preference for using reference letters and numbers rather than trying to find meaningful labels for his personality factors (Cattell, 1957). The analytic method, then, is limited by the current state of development of personality theory. A further drawback of a thoroughgoing analytic method of assessment is the practical consideration of economy of effort; the returns may be just as great, probably greater, in the first pilot assessments in a program, if we use an empirical or global strategy without trying to make explicit the underlying theoretical relationships. In addition, analytic assessments require a double inference and consequently the possibilities of error are increased;

[4] Knowledge of the results of this analysis did not improve the validity of the assessor's predictions, but this could have been caused by the assessors preferring to use a global rather than analytic strategy despite the analytic information which was supplied (Holt, 1958).

either the criterion analysis or the ratings of the candidate might be in errors. In the analytic strategy, however, there is at least the hope that the sources of these errors will be discovered and corrected, whereas the sources are masked in the nonmediated strategies.

The analytic strategy is applicable to any of the three purposes, selection, validation research or personality research, but its greatest potential is in the latter; in fact, if the results of assessment are to be of any value in increasing our understanding of personality, it is essential that the data be expressed in terms of basic personality constructs underlying the subject's behavior so that the scores and observations on the subjects may become meaningful. This applies both to naive empirical strategies such as factor analysis or blind item validation, and to global strategies in which the mediating constructs are not made explicit.

To sum up: we have argued that both the naive empirical and the global strategies are actually mediated by analytic personality constructs, but that it is not always necessary, or even possible to make those mediating variables explicit. This may apply both when the purpose of the assessment is validation of the techniques or the carrying out of an actual selection. But when the purpose is personality research, some explicit handling of the constructs is advisable. The concept of construct validation supports this requirement by merging the validation and the personality research orientations.

Each of the three strategies has its particular uses in assessment programs. Where mass screening is required, the empirical strategy is usually best, if possible; where the criterion situation is complex and unrepeatable, but familiar to the assessors, the global approach is to be preferred, and where the relevant personality theory has attained a sufficient level of development, the analytic strategy is indicated. Where none of the basic requirements are present—a repeatable and reliable criterion, familiarity of the criterion to the assessors, or appropriate personality theory—the assessors have to choose the strategy that seems best, although no strategy can really redeem such a hopeless situation. In general, personality assessors being what they are, they will prefer a largely intuitive approach, either analytic or global, as they did in the WOSB and OSS situations, but an increasing respect seems to be paid to the need for illuminating these intuitive methods by empirical analysis wherever possible.

Some Specific Issues in Assessment as a Method of Predicting

Clinical Versus Statistical Approaches

We have argued that there are occasions when intuitive methods of making predictions, i.e., "clinical" have their appropriate place. Statistical methods cannot be used where no prediction formula exists. But some personality assessors speak as if the clinical method is always to be preferred as it enables the assessor to be flexible in his use of the data in a way that is not possible with statistical techniques; for example, the clinician can give weight to obvious but rare and nonrepeatable factors in the subject's current situation which could not be validated empirically. Other advantages claimed for the clinical against the statistical approach are that it does not violate the essential unity of the subject's personality, and that it enables the

use of empathy and recipathy in making the predictions. (Actually these subjective clues could also be used as data by the statistician along with other more objective data.)

Other assessors regard clinical techniques as only a last resort. A number of advantages can be quoted for statistical prediction over clinical, most of which boil down to the fact that the statistician has a far more efficient memory and a larger attention span than the clinician; he can "remember" the relevant data at the appropriate time and combine them with other data in order to obtain optimal weightings for future predications.

And so we have, on the one hand, the efficient but rigid and inhuman statistical prediction, and on the other, the flexible and humane but inefficient clinical. Which one is more useful in personality assessment? There are several discussions of this question available (e.g., Cronbach, 1956; Meehl, 1954; Holt, 1958; Sarbin, Taft, and Bailey, in press; McArthur, 1954) so the points will not be elaborated fully here excepting in so far as they directly affect multiple personality assessment procedures.

The weight of the evidence clearly supports the accuracy of the statistical approach compared with the clinical. Meehl's notorious scoreboard (1954) recording the relative validity of clinical versus statistical prediction mounts grim evidence in favor of the latter. Holt (1958) criticizes Meehl's summary on the grounds that most of the studies pitted sophisticated, actuarial predictions against "naive clinical," while others (e.g., Wittman's) actually showed the superiority of "sophistical clinical" over naive clinical methods. But Meehl is quite clear about the rules of his contest: the rival methods start off with approximately the same objective and subjective data, although in some of the studies the clinician used additional subjective data. The important difference is that the reported statistical predictions were based on the naive empirical method of validation, while the clinical were either global or intuitively analytic. The statistical approaches were not concerned with the meaning of the correlations between the data and the criteria, although the use of cross-validation and statistical refinement meant that the empirical procedures were not as naive as it appeared, nor were they always uninformed by intervening personality constructs.

Holt pleads for the use of "sophisticated clinical methods," by which he means something similar to our analytic procedure, using intuition to make the final predictions. Among other things, he wants the clinicians to make preliminary studies of the criterion behavior, in order to analyze the requirements for success. Holt does not take the step of requiring validation of the individual clinicians; but this is necessary to match fully the two sides in the contest. He reports that the best judges, using global clinical techniques, reached prediction validities of up to 0.57, whereas statistical treatment of the tests—regular Rorschach scoring (validated and cross-validated) and the Strong Interest Psychiatrist key —resulted in virtually zero validities. But Holt's contest is unfair to the statistical side. His experiment was a half-hearted affair; no attempt was made to develop objective tests that would be appropriate to the selection problem at hand, as was done in the VA and the IPAR studies, and on Holt's own admission the Strong key

was validated a long time previously in an entirely different situation. Holt's report, as he himself indicates, does not provide us with a fair contest between sophisticated clinical and sophistical statistical approaches.

In recommending a sophisticated clinical approach, Holt argues that "there simply is no substitute for empirical study of the actual association between a type of predictive data and the criterion" (1958, p. 3). Despite this, the evidence that he presents on the value of objective criterion analysis for the assessor (i.e., validation) is not promising. The assessors at Menninger were provided with "manuals" embodying validation material on the interview, TAT, Rorschach, and other assessment techniques that had been used in an earlier assessment of psychiatrists at Menninger. Holt's conclusion about their value reads as follows: "Of the six, two proved worthless . . . ; the other four all showed more or less promise, but *there was none that yielded consistently significant validities regardless of who used it*" (1958, p. 8, italics ours). Evidently, the assessors would not, or could not, use the validation data which were provided for them.

We are thus reminded once more that validation includes validation of the specific assessors carrying out the specific assessment task. Nearly all reports of personality assessments offer evidence that the assessors differ considerably in their predictive skill. These differences are made up of two types of variation; variation due to differences in general ability to judge people (Taft, 1955), and interaction effects between the assessor and the type of judgment called for (Crow & Hammond, 1957). The reports on assessments offer the hint that the highest validities are achieved by assessors who have the most familiarity with the criterion situations and with the type of person who is successful in those situations; for example, in the CISSB assessments, the Board of Review consisting of experienced civil service administrators made more accurate predictions than did the original CISSB selection committee. In the former, the most valid predictions were made by the chairmen who were also civil service administrators.

Accurate assessments are most likely to occur where the assessor uses the in-group stereotypes which are also held by the criterion raters; they are able to "play their predictions by ear" without any need to make the double inference involved in analytic techniques. In support of this method of predicting we can quote the comparatively high validities found for ratings of the "likeableness" of the candidates in the Michigan, Menninger, and IPAR assessments. For example, in the latter, the assessors were mainly university professors, and it is therefore not surprising that their ratings of "personal soundness" correlated as high as 0.52 with ratings made of the candidates on this quality by their own departmental professors (Barron, 1954). All other things being equal, the best assessors for predicting existing criteria are those who are partially contaminated with the same experience, standards, and outlook as the criterion raters and can thus rely on a global strategy to make their predictions. (The most accurate assessors are also more accurate than the most accurate, cross-validated, tests.)

The validity of analytic methods is subject to the accuracy of the personality theory which the assessor

uses, but psychologists usually possess fairly stable postulates, based on the lore of their discipline rather than behavior-oriented empirical research, and these are not readily changed in the light of actual empirical data. This is probably the reason why some of the Menninger assessors did not improve their accuracy with the help of the empirically-based "manual." The difficulty can be seen clearly if we consider the findings of the Minnesota starvation studies (Kjenaas & Brozek, 1952) that the Rorschach indices of adjustment had a negative validity in predicting ratings of the subject's adjustment after starvation. Could a typical clinical psychologist bring himself to reverse completely his normal interpretation of the Rorschach in order to predict the subject's adjustment under the criterion conditions? Not unless he were able to find an intervening variable between the Rorschach and the criterion that would enable him to understand the connection within the framework of his existing theory of personality.

Our discussion of clinical versus statistical methods of assessment has concentrated on one aspect of the procedure, the prediction-making stage. The contrast between these two approaches can be made in connection with a whole chain of decisions that must be made in the course of assessment: these decisions include determining the acceptable criteria, scoring the criterion behavior, conducting the criterion analysis, determining the form of the tests and standard situations, observing and classifying the assessment behavior (i.e., scoring), combining the observations made by any single assessor into an assessment or prediction and combining the predictions made by different assessors. For example, should the individual assessments be combined subjectively by the chairman of an assessment board, by voting, or by averaging the individual predictions? Insufficient attention has been given to the relative merits of subjective and objective methods at each one of these stages.

The choice of method will depend on both the requirements and the over-all situation, including, sometimes, public relations considerations. The final selection of assessment techniques is likely to be a mixture of both subjective and objective, but the circumstances that will favor one or the other at any stage are rather vague, and the choice is usually made on subjective grounds, although it, too, could be made on the basis of objective, empirical investigation. In general, objective methods are to be preferred as far as possible as they maximize accuracy, but practical considerations of economy, convenience, and the limitations of the situation, dictate the wholesale use of subjective methods in personality assessment. These subjective methods may have high validity under favorable circumstances, and where the assessors are familiar with the criterion situation, clinical judgments may actually be more accurate than any objective methods are ever likely to be in predicting to criteria.

Conditional Variables in the Criterion

An old problem in evaluating the validity of prediction is set by variations in the criterion situation attributable to the surrounding conditions. For example, a prediction that a candidate will make a good officer may be invalidated through some contingency such as being posted to

a commanding officer with whom he is incompatible. But these conditional factors do not stand on their own; there is an interaction between the person and the condition. Thus, Officer A may have the type of personality (or background) that makes it likely that he will be posted to a commanding officer with whom he will be incompatible; if, for instance, Candidate A is Jewish, he is more likely to have a CO who behaves uncongenially than is another candidate of a similar personality who is not Jewish. Further, Officer A may perform his duties better than otherwise when he has an uncongenial CO, while Officer B may perform his duties worse under the same circumstances. In most assessments, no specific reference is made to such conditional factors and there is an implicit assumption of "given normal conditions" attached to the predictions. The OSS reports a validity of only 0.19 for all cases from Station S compared with 0.39 for only the cases who were given assignments that were consistent with the ones for which they were assessed.

A further condition that is often ignored in assessment is that of effluxion of time; the predictions are usually made on the assumption that the status of the candidate on the relevant variables will remain constant over time. At a more sophisticated level, trends towards change may be observed in the candidate together with potential but as yet unrealized capacities, and the assessment may extrapolate these into the future. But it is virtually impossible to take into account subsequent learning, maturation and deterioration in the assessment prediction.

In this connection, Cronbach and Gleser (1957) have proposed a useful distinction between fixed treatment (the same conditions for all successful candidates) and adaptive treatments varying according to the candidate. Evidently the treatment of the OSS selectees was fixed rather than adaptive, and the predictions should have taken this into account. Five different types of solutions are suggested below for the problem of conditional factors in these treatments, Solutions 1 and 3 being particularly appropriate to fixed treatments, and the other three to adaptive. (These represent an expansion of the three solutions proposed in Horst, 1941, ch. 5.)

1. Adjust the criterion ratings ex post facto according to the ease or difficulty presented to the candidate by the criterion conditions and the effects of these conditions on him over the relevant period of time. This adjustment requires an intuitive judgment that takes into account the interaction between the conditions and the candidate, and this can be done only by the rater making a further, independent assessment of the candidate. For the validation to carry conviction, it is necessary that the adjustment to the criterion rating be made independently of the assessment.

2. Make the predictions to the ideal possible conditions so that they represent the candidate's fullest potential; the criterion ratings can then be made in accordance with the same standards. In other words, both assessment and prediction attempt to hold conditions constant in the form that is considered to be optimal for the candidate's performance. The actual conditions applying at the time of assessment and during the criterion performance are unlikely to be optimal no matter how hard this state is sought, so that the use of this solution rests very heavily on intuition.

3. Predict to the average or modal conditions that have prevailed in the past with respect to the criterion situation, or which are expected to prevail in the future. This is the usual orientation in assessments based on empirical validation since the correlations on which the validities are based are in effect averages. The empirical strategy automatically takes into account the variations in conditions as well as their average effect and maximizes the prediction to these average conditions. The conditions, thus, influence the assessment only through their effects on

the criterion performance, without regard to their specific nature. It is practically impossible for a clinician to average all possible relevant conditions by an intuitive act, although it is common for a clinician to bear in mind the modal conditions which candidates face when these are prominent.

4. The future conditions may be predicted specifically for each candidate so that the interaction between the candidate and the criterion conditions may be anticipated in the assessment. The prediction to the future conditions may be made on the basis of inside knowledge of the treatment to be given to the candidates in the criterion situation, or by forecasting on general grounds the specific changes that will occur in the conditions before the criterion ratings are made. Such predictions must be intuitive rather than empirical, and, by the nature of the complexity of man's environment, all such intuitive predictions must fall well short of perfect validity. In some complex situations, in which the criterion performance is highly dependent on the conditions, the inability of the assessors to predict the specific conditions that will operate for any particular candidate may render the assessments completely invalid.

5. The predictions themselves can be made in terms of specific conditions: "if *X* conditions occur, then the candidate will be successful." In this endeavor, the recent proposal by Cattell (1957, pp. 426ff.) for a taxonomy of situations might eventually supply a list of standard situations to be considered in conditional predictions.

Solutions 4 and 5 are both specific conditional solutions which can take into account the effects of conditions that are external to the candidate, as well as intrinsic conditions such as maturation. They require both a knowledge of the criteria requirements and a correct assessment of the candidate, but the first type of conditional prediction emphasizes the criterion situation, and the second, the candidate. Both of these latter methods of meeting the problem of conditions are adaptable to taking into account multiple conditions and also "adaptive treatments" such as provisions for training that are tailor-made for the candidate. They hold out the possibility of making more exact predictions than can be made by the other three attempted solutions to the problem. This is one of the reasons why the global strategy, or slightly analytic versions of it, have been so often favored in selection assessment programs. But these conditional predictions are also the most difficult to make, and only the best judges of personality or the ones who are most experienced with the criteria conditions are able to make them accurately, and then only in appropriate situations.

The decision as to the appropriate solution to the problem of varying conditions is closely related to the choice of strategy. In the long run the choice is one between elegance and the practical limitations that are imposed on the possibilities of accuracy.

The Assumption of Safety in Numbers

Personality assessment programs rely on numbers to improve their validity in two directions: multiple tests and multiple assessments. We shall treat the evidence concerning these two points separately.

Multiple tests. Where the tests and other assessment measures are combined objectively, for example, in accordance with a multiple regression equation, even the most valid test can usually be improved upon by adding one or two further measures to it. It is often striking, however, how quickly the multiple Rs reach their ceiling; the common components of almost all available personality measures seem to be so high that we quickly exhaust the new elements that additional tests can bring to the predictions. The same applies when the combining of elements is carried out intuitively; even though the clinician may believe that the pieces of information about a candi-

date are independent of each other. It is doubtful whether a clinician can use more than a few pieces of data that are relatively independent, even if they can be found. Sarbin (1942), for instance, demonstrated that clinicians who were given a mass of data from which to predict the success of university students, gave most of the weight to two variables only.

Evidently, to give a clinician more than two or three pieces of data about an assessee is likely to be of little value. Some critics go even further, claiming that giving extra data actually reduces validities by confusing the allocation of subjective weights to the predictor variables, and by increasing the variability of the predictions, i.e., inducing the clinician to venture into making extreme judgments which increase the risk of making large errors. Kelly and Fiske claim (1950) that in the Michigan study validities declined as more data were given to the assessors. Holt challenges the accuracy of their interpretation of the findings (1958, p. 8), but even so there are other studies that suggest that more data do not always improve accuracy (e.g., Gage, 1953; Giedt, 1955; Kostlan, 1954; Soskin, 1954). In Giedt's study, for instance, the clinicians were able to make more valid predictions of mental patient's personalities from sound recordings than from sound movies.

But there are several studies affirming that, at least under some circumstances, more data do enable clinicians to improve their accuracy. We have already referred to Vernon's report (1950) that in the CISSB selections the Board of Review was able to improve on the assessment board's recommendations by combining these with their own interview impressions of the candidate. Increased validity with increased data is also reported for the California (Mackinnon, 1951), Chicago (Stern, Steiń, & Bloom 1956), and Menninger (Holt, 1958) assessments.

We must suspend our verdict on the value of multiple data at this stage. Evidently there are circumstances that can overcome the limitations on the ability of a single assessor to hold in mind data and to combine them. One suggestion worth testing is to combine data into subdecisions of an increasing degree of molarity, until the final molar decision is reached. This procedure can assist the clinician to consider all of the data in reaching his final decision and it is analogous to the use of structured schedules and rating forms that are used by interviewers to consolidate portions of the data as they go along. This technique as a general aid to clinical judgments seems worth experimenting with, although the danger must be avoided of giving too much weight to the data that are presented first. In this respect it would seem to be wise to seek out first the data that are believed to be the most valid.

Another way of handling the combining of data is to use several assessors, each responsible for one or two different techniques or areas of personality. This was the method adopted, for example, in the CISSB assessments. This proposal carries over to the general question of using multiple assessors and we shall consider it further below.

Multiple assessors. The practice of using more than one assessor in selection work is an old one; the assumption has been that the more assessors there are, the more ideas will be thrown into the pool and therefore the more thorough will be the marshalling of data. Where ratings are

pooled, it is also hoped that errors will cancel each other out. Very little experimental material is available on the relative value of group versus individual judgments in personality assessments, but evidence can be used from other work on other types of group performance (see Kelly & Thibaut, 1954; Klein, 1956, ch. 1; Argyle, 1957, ch. 5).

These findings suggest, among other things, that accuracy of judgments increases with the size of the group, but the optimum number in informal problem-solving groups is possibly five, since larger groups require formal structuring in order to ensure adequate communication of information; that compatible membership is important in problem-solving committees; that democratic groups produce more different ideas than individuals but fewer per person; that the quality of group decisions increases with an increase in the skill of the members; that groups are quicker at solving problems than individuals, although less economical in terms of man-minutes. However, these findings vary according to the type of task concerned, and before we can carry them over to personality assessment it is necessary to bear the type of task in mind.

Some of the questions that should be asked concerning group factors in personality assessment are: are group ratings more accurate than those of the individual members of the group; does group discussion by the assessors improve accuracy over pooled individual ratings; what is the relative value of means, modes, and medians as methods of pooling; the ideal size of committees; committee ratings versus averaging; authoritarian leadership of assessment committees versus democratic; should all of the committee members be given the same data; should both the observations and the interpretation be made by groups? These questions can be considered at three points in the assessment procedures: (*a*) in making subjective observations of the subjects; (*b*) in eliciting data from the subjects; and (*c*) in integrating and interpreting the data, and making the decision.

(*a*) At the observational level, we should expect that the pooled ratings of several observers would be more accurate than individual ratings, since pooling reduces the error variance, provided always that the individual judgments have some validity in the first place (*cf.* Kelley & Thibaut, 1954, p. 739).

(*b*) The value of the group interview versus individual interviews as a means of eliciting data is equivocal (see the discusssion in Oldfield, 1947). A recent study (Glaser, Schwarz, & Flanagan, 1958) on the selection of supervisors found that individual ratings based on group interviews by a panel of three were no more accurate than the ratings made by one interviewer per candidate. While it is true that the group situation may elicit a wider sample of behavior than an individual interview, it is more difficult for an interviewer to evaluate the significance of the group as the stimulus to which the candidate is responding. However, if the interviewing is conducted by the chairmen only, while the other assessors are simply observers, this may enable the assessors to make more unbiased judgments than when they are actually involved in the interviewing. This effect still remains to be tested empirically.

(*c*) The usual consideration of the value of group assessment versus individual deals with the integration of the available data. As in the case of group observations, pooled predic-

tions are more accurate than most or all of the individual predictions (Klugman, 1947; Luborsky & Holt, 1957; Travers, 1941). In one study (Smith, 1932) of assessing the qualities of a child on the basis of behavioral data, the accuracy increased with an increase up to 50 of the number of assessors whose ratings were pooled (there were only 50 assessors available).

Does discussion prior to assessment increase accuracy? The evidence on this suggests that it does not (Rusmore, 1944; Taylor, 1947; Kelly & Fiske, 1951; Oldfield, 1947). As Oldfield puts it: "Discussion of the merits of candidates merely amounts to a somewhat clumsy method of averaging the individual judgments of the members" (1947, p. 129). Whether discussion aids accuracy or not appears to depend on the quality of the persons who dominate the discussion either through their position in the group, their personality, or their professional standing. Discussion is justified particularly when there is an "expert" as chairman, who will actually make the final decision in an autocratic manner, but who calls on the other members of the panel to give him the benefit of their opinions. An "expert" is defined, for this purpose, as a person who is experienced both in assessment and in the criterion situations.

Kelly and Fiske are quite pessimistic regarding the use of multiple assessors. "Until some of the major sources of error in predictions are eliminated, the replications of assessors and the use of staff conferences hardly seems justified for this type of prediction" (1951, p. 178). This conclusion is too sweeping. As we can see from Table 1, both pooled and committee (discussion) ratings have justified themselves in some studies.

Let us conclude this section on the "safety in numbers" assumption with a proposal to combine the advantages of both multiple techniques and multiple assessors. The suggestion is that each assessor be given a limited amount of information on which to base his assessment judgments about the candidates, each assessor to receive different information. The assessments will then be pooled arithmetically. The information supplied may be objective or subjective, atomistic or molar, and may range from one item of life-history, or a test result, to a projective test protocol, an interview or the observation of behavior in a miniature situation. This procedure would enable a vast amount of data to be integrated without problems of weighting since unit weights for each assessor's contribution would be adequate—this would be analogous to an inventory that gives unit weight to each item. With adequate organization of the assessment program, this would permit several assessors to contribute to the final assessment so that different viewpoints and personality theories can be represented. This approach seems to be at least worth experimenting with.

Even if it is found that increased numbers of assessors increases perceptibility the accuracy of the assessments, there is still a fine calculus of cost in human time and effort to be computed. The decision to augment the panel with additional assessors is a function, among other things, of the gradient of diminishing returns, the ability of available extra assessors, the cost of using them, the effects on the candidates, and the desire to allow executives in the institution to participate in the assessment. The proposal made above of having many assessors, who contribute small pieces

of information, may make it possible to conduct multiple assessments comparatively cheaply.

Summary

Multiple personality assessment procedures have been analyzed with respect to their primary purpose and the validation strategy used. Problems that arise in the attempt to use personality assessment for selection were discussed with respect to the problem of clinical versus statistical predictions, the problem of conditional factors that affect the criteria, and the value of using multiple tests and more than one assessor.

Some recommendations:

1. Use objective techniques as far as possible for analyzing the criterion, for scoring tests, and for making predictions.

2. Give careful consideration to requirements of the criterion and make empirical studies of the link between these requirements and both the test behavior and the criterion behavior. This is a step in construct validation.

3. As a preliminary to the above type of criterion analysis, nonmediate empirical or clinical methods of prediction may be used.

4. Subjective observations should be made by several observers whose opinions should be pooled arithmetically.

5. The assessors should be selected for proven ability to make accurate judgments in the assessment situation, i.e., they should be validated.

6. The assessors should be familiar with the criterion situation, and should take this situation into account when they make the predictions.

7. Each assessor should be given no more than two or three units of information; there should be a large number of assessors whose predictions are pooled arithmetically, and without discussion.

8. In selection assessments, if committee decisions are desired, the assessors who are particularly well-experienced in the criterion situation should be given special influence in forming the final decisions, provided they have been shown to possess good ability to judge persons.

REFERENCES

Barron, F. *Personal soundness in university graduate students.* Univer. of California, 1954.

Cattell, R. B. *Personality and motivation structure and measurement.* New York: World Book, 1957.

Cronbach, L. J. Assessment of individual differences. In *Ann. rev. of psych.*, Vol. 7. Stanford: Annual Reviews, 1956.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.

Cronbach, L. J., & Gleser, Goldine C. *Psychological tests and personnel decisions.* Univer. of Illinois, 1957.

Crow, W. J., & Hammond, K. R. The generality of accuracy and response sets in interpersonal perception. *J. abnorm. soc. Psychol.*, 1957, **54**, 384–390.

Eysenck, H. J. *Dimensions of personality.* London: Kegan Paul, 1957.

Farago, O., & Gittler, L. F. (Eds.) *German psychological warfare: Survey and bibliography.* New York: Comm. for Nat. Morale, 1941.

Gage, N. L. Explorations in the understanding of others. *Educ. psychol. Measmt.*, 1953, **13**, 14–26.

Giedt, F. H. Comparison of visual, content, and auditory cues in interviewing. *J. consult. Psychol.*, 1955, **19**, 407–416.

Glaser, R., Schwarz, P. A., & Flanagan, J. C. The contribution of interview and situational performance procedures to the selection of supervisory personnel. *J. appl. Psychol.*, 1958, **42**, 60–73.

Gough, H. G. The construction of a personality scale to predict scholastic achievement. *J. appl. Psychol.*, 1953, **37**, 361–366.

Harris, H. *The group approach to leadership testing.* London: Kegan Paul, 1959.

Holt, R. R. Clinical and statistical predic-

tion: A reformulation and some new data. *J. abnorm. soc. Psychol.*, 1958, **56**, 1–12.

HOLT, R. R., & LUBORSKY, L. *Personality patterns of psychiatrists.* New York: Basic Books, 1958.

HORST, P. *The prediction of personal adjustment.* Soc. Sci. Res. Counc. Bull. No. 48, 1941.

KELLY, H. H., & THIBAUT, J. W. Experimental studies of group problem-solving and process. In Lindzey, G. *Handbook of social psychology.* Vol. 2. Cambridge; Mass: Addison-Wesley, 1954.

KELLY, E. L. Multiple criteria of medical education and their implications for selection. In *The appraisal of applicants to medical schools.* Rep. of Fourth Teaching Instit. Evanston: Assoc. of Amer. Med. Coll., 1957.

KELLY, E. L., & FISKE, D. W. The prediction of success in the V. A. training program in clinical psychology. *Amer. Psychologist*, 1950, **4**, 395–406.

KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology.* Univer. of Michigan, 1951.

KJENAAS, NANCY K., & BROZEK, J. Personality in experimental semi-starvation. *Psychosom. Med.*, 1952, **14**, 115–128.

KLUGMAN, S. F. Group and individual judgments for anticipated events. *J. soc. Psychol.*, 1947, **26**, 21–28.

KOSTLAN, A. A method for the empirical study of psychodiagnosis. *J. consult. Psychol.*, 1954, **18**, 83–88.

LUBORSKY, L., & HOLT, R. F. The selection of candidates for psychoanalytic training. *J. clin. exp. Psychopath.*, 1957, **18**, 166–176.

MCARTHUR, C. Analyzing the clinical process. *J. counsel. Psychol.*, 1954, **1**, 203–207.

MACKINNON, D. W. The effects of increased observation upon the accuracy of prediction. *Amer. Psychologist*, 1951, **6**, 311. (Abstract)

MEEHL, P. E. *Clinical versus statistical prediction.* Univer. of Minnesota, 1954.

MORRIS, B. S. Officer selection in the British Army, 1942–1945. *Occup. Psych.*, 1949, **23**, 219–234.

MURRAY, H. A. *Explorations in personality.* New York: Oxford, 1938.

OLDFIELD, R. S. *The psychology of the interview.* (3rd ed.) London: Methuen, 1947.

O.S.S. *Assessment of men.* New York: Rinehart, 1948.

RUSMORE, J. T. An experimental comparison of the composite and jury methods of obtaining group judgments. Unpublished doctoral dissertation, Univer. of California, 1944.

SARBIN, T. R. A contribution to the study of actuarial and individual methods of prediction. *Amer. J. Sociol.*, 1942, **48**, 593–602.

SARBIN, T. R., TAFT, R., & BAILEY, D. *Clinical inference and cognitive theory.* New York: Rinehart, in press.

SMITH, M. Group judgment in the field of personality traits. *J. exp. Psychol.*, 1932, **14** 562–565.

SOSKIN, W. F. Bias in post-diction from projective tests. *J. abnorm. soc. Psychol.*, 1954, **49**, 69–74.

STERN, G. G., STEIN, M. I., & BLOOM, B. S. *Methods in personality assessment.* Glencoe, Ill.: Free Press, 1956.

TAFT, R. The ability to judge people. *Psychol. Bull.*, 1955, **51**, 1–23.

TAYLOR, D. W. The analysis of predictions of delinquency based on case studies. *J. abnorm. soc. Psychol.*, 1947, **42**, 45–56.

THORNDIKE, R. L. *Personnel selection.* New York: Wiley, 1949.

TRAVERS, R. M. W. A study in judging the opinions of groups. *Arch. Psychol., N. Y.*, 1941, No. 266, 1–73.

VERNON, P. E. The validation of civil service selection board procedures. *Occup. Psychol.*, 1950, **24**, 75–95.

PSYCHOLOGICAL BULLETIN
VOL. 56, NO. 5, 1959

THE CONCEPT OF PROJECTION: A REVIEW

BERNARD I. MURSTEIN
University of Portland
AND
RONALD S. PRYER
State Colony and Training School, Pineville, Louisiana

The word "projection" stems from the Latin verb *projectus*, meaning "to cast forward" (Bell, 1948). In the field of personality one would be hard pressed to find a concept so capable of multiple interpretation and so varied in meaning as the concept of projection. Surely, this concept has had more interpretations than the smile of Mona Lisa. Sears described the situation thus: "Probably the most inadequately defined term in all psychoanalytic theory is projection" (Sears, 1943, p. 121). Murray has said "If 'projection' means everything it means nothing" (Murray, 1951, p. 13).

Just how many kinds of projection are there? It is frankly impossible to describe them all. The best one may hope for is to group them into more or less distinct categories. This paper, then, will concern itself with attempting to describe various "types" of projection, their origins, and the research undertaken with regard to them.

Only the various kinds of projection subjected to experimental investigation will be dealt with in detail. Because they have resulted in little research, the conceptual frameworks of Bellak (1950, 1954, 1956), Murray (1938, 1951), Van Lennep (1951, 1957), Rapaport (1942, 1945, 1952), Schachtel (1950), and Goss (1957) will not be considered.

The research literature seems to indicate some four possible categorizations of the concept of projection: "classical," "attributive," "autistic," and "rationalized" projection. Under the category "miscellaneous" we will review those studies in which projection was undefined, or in which several different concepts were simultaneously investigated.

Classical Projection

The concept of projection was known several centuries before the appearance of Freud. Thus Thomas a Kempis stated, "What a man is inwardly that he will see outwardly" (Cattell, 1951). The Malleus Maleficarum, written in the Middle Ages, gives a clear example of projection.

> For fancy or imagination is as it were the treasury of ideas received through the senses. And through this it happens that devils so stir up the inner perceptions, that is the power of conserving images, that they appear to be a new impression at that moment received from exterior things (Zilboorg, 1935, p. 54).

At the close of the 19th century, Freud gave the following definition of projection: "The psyche develops the neurosis of anxiety when it feels itself unequal to the task of mastering (sexual) excitation arising endogenously. That is to say it acts as if it had projected this excitation into the outer world" (Bellak, 1944, p. 353). With this definition, the use of the concept soon became quite widespread.

The view of "classical" projection most often held currently by many psychologists is: "A situation in which the ego feels threatened is likely to result in the ego's refusing to acknowledge the trait and in the sub-

sequent attribution of the trait to the outside world" (Murstein, 1956, p. 418). Adherents to this view are many, though they differ as to the wording used. Only a representative sampling is listed (Healy, 1930; Hoffman, 1935; Jelgersma, 1926; Kaufman, 1934; Knight, 1940; Schafer, 1954; Schaffer, 1945; Symonds, 1949; Warren, 1934). The possibility of the projection of objectively favorable traits has been mentioned by Muhl (1943), Hoop (1924), and Janet (1947).

Attributive Projection

"Attributive projection" has been described by many psychologists, including Cameron (1951). A recent definition is: "The ascribing of one's own motivations, feelings, and behavior to other persons" (Murstein, 1957b). It is perhaps the most popular of the uses of projection today in the field of personality. The concept's popularity rests on its broadness; i.e., unlike "classical" projection there is no concern with the *S*'s unconscious, or self-concept. It is often sufficient merely to note that there is a correlation between some characteristic of the subject and some statement or prediction he makes concerning other persons. Weiss points out that "the term projection in current usage, refers to every kind of externalization, particularly to every process in which ideas, impulses, or qualities belonging to oneself are imputed to others" (1947, p. 358). Nevertheless, the concept has strong support (Cameron: 1947, 1951; Dymond, 1950; Munn, 1946). Even Freud, in *Totem and Taboo* said (as noted earlier by Bellak, 1956):

> But projection was not created for the purpose of defense; it also occurs where there is no conflict. The projection outwards of internal perceptions is a primitive mechanism, to which, for instance, our sense perceptions are subject, and which therefore normally plays a very large part in determining the form taken by our external world. Under conditions whose nature has not yet been sufficiently established, internal perceptions of emotional and thought processes can be projected outwards in the same way as sense perceptions, they are thus employed for building up the external world, though they should by rights remain part of the *internal* world (Freud, 1955, p. 64).

Freud's statements are, thus, amenable to an attributive definition of projection while comprehending projection as an organizing aspect of perception.

Horney also described a nondefensive kind of projection which, in part, is "not essentially different from the tendency to assume naively that others feel or react in the same manner as we ourselves do" (1939, p. 26). An example of "naive" projection is mentioned by Baldwin (1955) in describing the child who in a fit of anger threatens to maim, murder, and demolish another child, and as a crowning imprecation, threatens not to invite his adversary to his birthday party.

Autistic Projection

Perception which is strongly influenced by the needs of an individual, in that the figural aspects of the perceived object are modified so as to be consistent with the need, may be referred to as "autistic" projection. Murphy (1947, pp. 338 ff.) wrote "so wherever our needs differ we literally *see* differently. Much of the process of individual perception depends upon the force of past wants, the person's need to disentangle and restructure in terms of the situations with which he has had to cope." Sears (1943, p. 121) has said, "it may be said in general, that the presence of a need or drive provides the antecedent condition for the perception of objects related to that need or drive."

Rationalized Projection

This type of projection is similar to "classical" projection in that the projective *process* is held to be unconscious. The projector, however, is conscious of his behavior. He attempts to justify it by inventing a rationale. Thus, the person caught buying on the "black market" says in self-justification, "everybody else is doing it." Here, the attempt is to convert neurotic anxiety about doing something wrong into objective anxiety about not getting enough to eat (Murstein: 1956, 1957b). Baldwin (1955) states:

> Finding a child disobedient and unloving justifies resentment. This introduces us to another sort of defense against guilt—namely to justify the rejection so that it no longer evokes guilt feelings. This attribution of blame to the people toward whom we feel hostile is a defense mechanism called "projection" (Baldwin, 1955, p. 498).

Among other supporters of this kind of projection are Allport (1939), (complementary projection), Fenichel (1945), The Psychiatric Dictionary (1940), and Piaget (1926), (projection de réciproque). Van Lennep (1951) has referred to this occurrence not as projection, but as its correlate. Thus, if a person is *frightened*, in order to project, he would have to see the *object as frightened*. If he sees the other as the cause of his fright and thus *frightening*, he is perceiving the correlate of his fright, but he is not projecting. Again, however, we find Freud in *Totem and Taboo* giving plausibility to a rationalizing kind of projection. He states:

> It cannot be disputed that this process of projection, which turns a dead man into a malignant enemy, is able to find support in any real acts of hostility on his part that may be recollected and felt as a grudge against him: his severity, his love of power, his unfairness, or whatever else may form the background of even the tenderest of human relationships (1955, p. 63).

At this point it may be of value to examine the experimental findings with regard to the aforementioned kinds of projection.

Experimental Findings

Classical Projection

Sears (1936) was among the first of the psychologists to utilize a quantitative index of projection. He studied male college students' possession of the traits of "stinginess," "obstinacy," "disorderliness," and "bashfulness," by the pooled rating method. Projection was measured as follows:

> (1) The degree to which each individual demonstrated a given trait was determined by averaging the combined ratings assigned to him on that trait by other members of his house. (2) The amount of a given trait attributed to others was obtained by averaging the ratings a given individual assigned to the other members (Sears, 1936, p. 153).

Projection would have occurred if, for example, stingy, noninsightful persons saw more stinginess in others than did the total group as a whole. The results, however, did not support a Freudian ("classical") concept of projection, in that projection occurred for both acceptable and nonacceptable (reprehensible) traits. Moreover, in the group possessing insight, a negative correlation was found between the strength of the trait and the amount attributed to others. Sears called this occurrence "contrast formation." Its effect was said to be opposite to that of projection.

One might have wished that "projection" and "insight" had not been treated dichotomously. The crudeness of measure due to the lack of a specific quantitative score other than "projected" or "did not project" defies any attempt at extensive analysis. Further, the appearance of a spurious element in the data, a fact noted by Rokeach (1945), and Calvin

and Holtzman (1953), places Sears' findings in doubt. The difficulty lies in the fact that projection as defined by Sears was a function of the group rating (G), and the difference between the group- and self-ratings (G–S). Since G appeared in both (G) and (G–S), a spurious correlation would be expected between (G) and (G–S), even if no actual psychological relationship existed. Hence, the effect of G should have been partialled out, something which was not done.

Zucker (1952) measured projection by summing the number of items in which an individual (college student) said that other people behaved in a certain way, but he, the student, did not. The approach seems questionable. No mention was made of the *actual possession* by the *S*s of the trait in question. Hence, *deviates from the population norm* would have been considered as projectors through the use of this methodology. Under these circumstances Zucker's findings that "high projectors" had more ideas of reference and were more ascendant on the Allport Ascendant-Submissive scale than "low projectors," seems meaningless.

Zimmer (1955) presented students with three photos and had them choose the ones liked best and least. The selected photos were then rated on a 7-point scale as to possession of some 25 traits. The *S*s then rated themselves on a scale containing the same traits. Finally, measures of conflict with regard to each of these traits were obtained from a word association test.

Zimmer's hypotheses, that acceptable personality characteristics are projected onto liked individuals, and unacceptable characteristics are projected onto disliked persons, and that the strength of projection is a function of the degree of conflict, were all substantiated. Similarly, Lundy and Berkowitz (1957) found that students whose attitudes were influenced by peers tended to perceive themselves as more similar to these peers than students who were negative to peer influence.

Norman, in a series of experiments with co-workers Ainsworth (1954), and Leiding (1956), investigated the relationship of such variables as "projection," "empathy," and "reality," and found the correlations given in Table 1.

These results apparently indicate that the various kinds of projection are detrimental to the accurate perception of others, but they have been shown to be an artifact of the procedure by Cronbach (1955), Gage and Cronbach (1955), and Murstein (1957a, 1957b). The criticism by Murstein (1957a) of their procedure is based on their definitions of projection, and is as follows:

Norman and Ainsworth Projection: "A" says that he does not possess a certain trait (he answers "no" to one of the questions on the GAMIN). "A" says that other students of his age and sex do possess that trait. Of the remaining members of the group (college students) from whence "A" stems, 51% or more say that they do *not* possess the trait in question.

Norman and Leiding, Projection A: This measure of projection may be more readily understood by means of an alphabetical shorthand used to describe each step (a, b, c, d, $\cdots$).

(a) "A" rates "B" as he thinks "B" would rate himself.
(b) "A" rates himself ("A") as he thinks "B" would rate him.
(c) "A" rates "B" as he ("A") sees "B."
(d) "A" rates himself ("A").
Projection = ($a-c$) plus ($d-b$)

Norman and Leiding, Projection B:

(e) "A" says most other people will answer a given question in a certain way.
(f) "A" answers in the same way in judging himself ("A") as he predicted that most other people would answer. A "pro-

TABLE 1

INTERCORRELATIONS OF "PROJECTION," "EMPATHY," AND "REALITY" TAKEN FROM THE NORMAN AND AINSWORTH (1954) AND NORMAN AND LEIDING (1956) STUDIES

Authors	Variables	r
Norman and Ainsworth	Projection vs. Reality	−.41
Norman and Ainsworth	Projection vs. Empathy	−.65
Norman and Leiding	Projection A vs. Refined Empathy	−.41
Norman and Leiding	Projection B vs. Raw Empathy	.86
Norman and Leiding	Projection B vs. Refined Empathy	−.69

jection" point is achieved each time e and f occur together for the same question. Thus, Projection $= \Sigma(e, f)$.

The projection score of Norman and Ainsworth may be objected to (*a*) because of the use of the subject's self-rating as an objective criterion of whether or not that subject possessed a given trait. This appears hazardous, particularly in a study measuring projection. Since the criterion for the group was 51% or more, if only 2% of the *S*s projected, the item would have been placed incorrectly and thus distorted the measure of projection; (*b*) because of the absence of any reliability coefficients for the various judgments; (*c*) because some of the traits on the GAMIN have moderately high intercorrelations, thus lending a degree of spuriousness to the results; (*d*) because in correlating "Projection" and "Empathy" there is a common component (others say they possess the trait) in both variables which spuriously inflates the resulting correlation.

Norman and Leiding used two other measures of projection which have been labeled by the present authors as Projection A and B. Using the aforementioned alphabetical shorthand for operations, their correlation of Projection with Refined Empathy (Raw Empathy minus Projection), for which a correlation of −.47 was reported, may be described as follows:

Projection vs. *Refined Empathy* (Raw Empathy − Projection)
$[(a - c) + (d - b)]$ vs. $[(a - g) + (b - h)] - [(a - c) + (d - b)]$

Similarly for Projection B, the correlation of Projection with Raw Empathy ($r = .86$) may be described as follows:

$$\frac{\text{Projection}}{\Sigma(e, f)} \text{ vs. } \frac{\text{Raw Empathy}}{\Sigma(e, i)}.$$

We have omitted the descriptions of g, h, and i in the interest of space since they are not crucial to our discussion.

In the first of these two correlations, the position of common components is such as to insure the fact that as the projection score increases, the refined empathy score must decrease due to the common component $[(a-c)+(d-b)]$. Hence, a spurious negative correlation is quite expected. In the second of the correlations, there is again an occurrence of identical components, but, here, since no component is subtracted, the spurious correlation which results is positive.

Murstein (1956), by means of pooled ranks, selected four personality groupings (hostile-insightful, hostile-noninsightful, friendly-insightful, and friendly-noninsightful). Using the Rorschach Hostility Scale as a measure of the hostile content of the Rorschach, he found that hostile, insightful, people projected more hostility than any other grouping. In a dynamic ego-threatening situation, however, the hostile, noninsightful,

group manifested "classical projection" as expected, but, surprisingly, the friendly, insightful, group also reacted strongly to threat, by distorting the examiner's behavior in perceiving him as hostile. Thus, the results did not wholly jibe with a "classical" conception of projection but were more amenable to analysis in a phenomenological frame of reference.

Both the hostile, noninsightful, and friendly, insightful, groups perceived themselves as friendly. Objectively speaking, the members of the friendly, insightful, group were correct, the members of the hostile, noninsightful, group were in error. What was important, however, was the way in which each individual perceived himself. The experimental findings are consistent with the belief of Lecky, that "any value . . . which is inconsistent with the individual's valuation of himself cannot be assimilated; it meets with resistance and is likely, unless a general reorganization takes place, to be rejected" (Lecky, 1951, p. 153).

In a quite similar design, Page and Markowitz (1956) obtained a defensive and nondefensive population from their responses to MMPI items. Again, a test of defensiveness based on receiving critical comments with regard to performance on a pseudo "intelligence test" did not significantly differentiate the defensive and nondefensive persons. Under personal ego-threat, the defensive persons projected more hostility on the examiner than did the nondefensive persons. The nondefensive persons did, however, show a nonsignificant tendency toward projection.

These two experiments seem to support the concept of "classical" projection in indicating that defensive persons do project under ego-threatening situations, though they may not do so on projective and paper-and-pencil tests. The failure of Page and Markowitz to find significant projection on the part of the nondefensive group may have been due to their different method of measuring defensiveness, as well as to the fact that their group consisted mainly of women while Murstein's consisted of men. Lastly, the sex of their examiner was not indicated, and may have been a woman, while in Murstein's study the examiner was a man. It seems quite plausible to believe that the sex of the subject, as well as that of the examiner, influenced the kind of response elicited.

It is difficult to draw conclusions from the aforementioned "classical" projection studies due to the varying methods of measuring the concept. Murstein's work pointed out the influence of background factors for the manifestation of projection, as well as the occurrence of projection with both "friendly" and "hostile" persons. It would appear, therefore, that the mechanism of projection is best understood as *a means of attaining self-consistency* rather than solely as a defense mechanism. Within this framework, the denial and subsequent projection of *favorable* traits may be readily understood. The "tough guy" may vigorously repress any inclination of "humanity" towards others since such behavior would be inconsistent with his self-perceived role.

Attributive Projection

By far the greatest amount of research has been undertaken within the confines of this concept. Sears (1937) found a correlation of .64 between an inventory measuring self-criticism and one tapping ideas of reference. Apparently, self-critical persons perceived others as also being critical of them. Thomsen (1941) found that in the 1940 presidential election, the majority of persons

sampled were of the opinion that the candidate whom they favored would win the election. Wallen (1941) sampled 85% ($N=237$) of the students in a small residential college asking them to estimate the percentage of students in the college who held certain opinions on each of three current issues (war entry, draft, St. Lawrence seaway), and in addition to state their own views. A significant proportion overestimated in the direction of their own opinions.

Goldings (1954) had 20 Harvard students rate pictures of college students on a 10-point scale of happiness, and then rate themselves on a 6-point scale. Some relationship was found between the self-rating and ratings of the pictures. A tendency was found, however, for individuals who avowed extreme happiness, to attribute less happiness to others. This effect may have been a function of regression to the mean, low reliability, and the "ceiling effect" of "happy persons." It will be readily apparent that no "extremely happy" person (rated 10 on a 10-point scale) could perceive anyone as "happier" than himself; only "equally happy" or "less happy" persons might be perceived, making for a negative correlation between self and others for such persons.

Mintz (1956) found that children who had just seen the film "Peter Pan" tended to perceive Peter Pan's age as close to their own. The correlation for boys was .34, and for girls .68, both significant.

Halpern (1955) had 38 student nurses take the GAMIN personality inventory, and had them predict the responses of other nurses in their immediate subgroup. Predictions for referents having self scores similar to their own were significantly more accurate than were predictions for dissimilar persons. The r for similarity vs. predictive accuracy was .84. Similar findings are reported by Suchman (1956), and Wittich (1956). Alfert (1958) found that when item responses were socially desirable, "attributive" projection (Alfert used the term "assumed similarity") increased for traits congruent with the perceiver's ideal more than items involving ideal-discrepant traits. Though the persons whose responses were predicted were comparative strangers to the predictor, they were probably of similar education and socioeconomic status. Thus, Alfert's findings that more items involving traits congruent with the *S*'s ideal self were attributed to these "strangers" than were said to be possessed by the *S* himself seems understandable. One may wonder whether such favorable perceptions of others would have occurred among persons of another social or economic group.

Fiedler (1951, 1952, 1953) found that persons liked were assumed to be more similar to the self than persons disliked. Good therapists, for example, showed a greater tendency to assume similarity unjustifiably between patients and themselves than was the case for poor therapists. These studies seem to indicate the independence of the concept of "attributive" projection from defensive behavior.

The above findings clearly demonstrate that people often predict that other persons, usually of *a similar occupational or social group*, hold many views similar to their own. Moreover, these predictions are often better than chance. This finding, however, may be subsidiary to the fact that many of the experiments involved situations where there was little ego-threat. In addition, most of the studies involved rather homogeneous groups where, due to the similar personality patterns of the members, the predictions of *accurate predictors* would undoubtedly show cor-

relations with their own self responses, thus automatically making them "projectors."

Lundy (1956) has shown that when the center of attention is on the self, projection is more likely to occur than when it is on the other person. When, however, an ego-involving situation is at stake, it might be expected that individuals would be more careful in making distinctions between themselves and others. Thus, Rokeach (1945), studying the ego-involving attribute of "beauty" among a female population at Brooklyn College, found a nonsignificant correlation of −.08 between self-ratings and ratings ascribed to other females.

The role of self-involvement as distinct from insight is sharply illustrated in a study by Weingarten (1949). She had a group of 74 college students write autobiographies of themselves. Two judges used the autobiographies to rate the subjects on "tension" and "insight" with regard to self, family, and the nonfamilial social environment. The *Ss* also were given a series of 75 statements describing behavioral incidents and asked to interpret the psychological import of these incidents. Projection was measured by the correlation between the judges' ratings of the subject's tension as expressed in his autobiography, and his interpretation of the behavioral incident inventory. The results appear in Table 2. From these correlations it is apparent that "attributive" projection is not necessarily related to insight. Where the "self" was involved, both high- and low-insightful groups tended to project. Again, it appears helpful to view this result as signifying a "self"-enhancing tendency for most persons regardless of their insight into their own personalities. By perceiving behavior in the behavioral incident inventory as similar to their own, as described in their autobiographies, the *Ss* tended to justify their own behavior as similar to the norm.

TABLE 2

"PROJECTION" CORRELATIONS FOR HIGH- AND LOW-INSIGHT GROUPS FOR SELF, FAMILY, AND SOCIAL ENVIRONMENT (WEINGARTEN, 1949)

Variable	High-Insight	Low-Insight
	(N=24)	(N=33)
	r	r
Self	.41	.38
Family	.00	.40
Social Environment	.15	.21

A welcome relief from the omnipresent college population is afforded in a study by Friedman (1955). He used three groups, 16 normals, 16 psychoneurotics, and 16 paranoid schizophrenics, all of whom *Q*-sorted items pertaining to their "phenomenological" self and their "ideal" self. In addition, they were presented with five TAT cards and their resulting themes were rated as to the presence of "projected" self.

The findings revealed significant relationships between "phenomenological" self sorts, and "projected" self on the TAT, for the normal and the neurotic groups, but not for the psychotic one. Considering the "ideal" self sort vs. "projected" self on the TAT, only the correlation of the normal group proved to be significant. The correlations revealed no significant differences between normal and neurotic population behavior concerning the "phenomenological" self and TAT protocols. Both groups showed a tendency to reveal something about themselves on the TAT, probably similar to that given on the *Q* sort. While the normal group projected their "ideals"

on the TAT cards, however, the neurotic group did not. Once again, the manifestation of projection seemed to be a function of self concept. The more anxious neurotic group was less consistent in its manifestation of "self."

Some writers have attempted to improve the measurement of projection by more quantitative methodologies than the earlier writers used. Rokeach (1945), for example, in his study of beauty, designated as insightful those subjects who in the beauty group slightly overestimated themselves, and in the homely group slightly underestimated themselves. His reasoning was that regression to the mean made the group judgments a slight underestimation of the true score of the individual. But, regression to the mean is a function of reliability; the greater the reliability, the less the regression toward the mean. Since no reliability coefficients were obtained with regard to the group judgments, Rokeach's method seems questionable.

Bender and Hastorf, in a series of experiments, sought an exact quantitative approach to projection (1950, 1952, 1953). A typical experiment (1952) required 50 students to fill out an adjustment scale for themselves and for the way they thought their friends would rate themselves. The variables were defined as follows:

Projection: The total item-by-item deviation of the forecaster's own responses from his predictions for an associate. *Empathy:* The deviation between the forecaster's prediction for his associate and his associate's self-rating. *Refined Empathy:* Empathy minus Projection. *Similarity:* The deviation of a forecaster's self-rating from his associate's self-rating. The correlations obtained, are shown in Table 3.

TABLE 3

CORRELATION OF "PROJECTION" WITH "SIMILARITY," "EMPATHY," AND "REFINED EMPATHY" (BENDER AND HASTORF, 1952)

Variables	r
Projection vs. Similarity	.32
Projection vs. Empathy	.37
Projection vs. Refined Empathy	−.58

Similarly, Cowden (1955) in a study of married couples, also found a substantial correlation between *Projection* and *Empathy.*

The results of both studies seem to imply that the more similar the personalities of two individuals, the more likely is projection apt to occur when one individual attempts to evaluate the personality of the other. The ability to see an individual as he sees himself (empathy) is related to the tendency toward projection. If, however, the projection component is removed from the score of Empathy, then a negative relationship exists between these two variables.

The conclusions, however, are an artifact of the statistical procedure. A contributing factor to the spurious relationship between projection and empathy as defined above is the fact that each of these scores had an identical component (prediction for an associate). The other scores illustrate similar common components when they are correlated with each other. An objection might also be made on purely logical grounds to the assumption that projection has occurred if a person attributes a trait to another which he possesses himself (Murstein, 1957b).

In a more recent article, Hastorf, Bender, and Weintraub (1955) brought out another difficulty with their earlier work. People who have a tendency to give end scale responses (almost always, almost never), and who also accurately predict midscale personalities, assure themselves of high Refined Empathy scores. The reason for this is that Refined Em-

pathy is equal to Raw Empathy minus Projection. Hence, the more dissimilar the personalities, the smaller the projection component subtracted, and the larger the Refined Empathy score.

By way of confirmation, the authors reported a rho coefficient of .47 between the tendency to give end scale scores and the Refined Empathy score. Conversely, a man who used the midpoint of the scale most frequently in his predictions for others, and also used midpoint prediction for himself, automatically became a projector. In short, projection under these circumstances was a function of response habit rather than any aspect of personality or cognition.

The independence of response habit and predictive efficiency is indicated in a study by Leventhal (1957). On the basis of the number of descriptions used to describe their friends, he derived groups of perceptually "simple" and "complex" judges. "Simple" judges were more apt to assume similarity between themselves and others in their predictions than were "complex" judges, who tended to differentiate between themselves and other persons. Nevertheless, both groups were about equal in predictive accuracy. In a study by Crow (1957b), medical students subjected to training in interpersonal relationships manifested greater complexity in their predictions of patients' MMPI profiles (i.e., the variance of their predictions increased as compared to their pretraining predictive variance), than they had prior to this training. Nevertheless, their accuracy dropped as a result of this training. Apparently, by overstressing the "individual," the basic similarity of persons or at least of their answering habits may be underestimated.

In a study involving a plethora of discrepancy scores, Fabian (1954) defined projection as the rating of "others" by a subject, minus the mean rating of the subject by the group. Insight was measured by the self-rating minus the mean rating of *S* by the group. Again, because of the component common to both measures (mean rating of *S* by the group), a (spurious) positive correlation was reported for projection and insight.

Bieri and his co-workers (1953, 1955a, 1955b) have been interested in the relationship of the interaction of persons and changes in perception. In an early study, Bieri (1953) found that after interaction for some 20 minutes, college students' "assimilative" projection scores (predicting the same responses for another as one fills out for oneself) tended to increase significantly. The test used to measure these predictions was the Rosenzweig P-F scale. Not surprising was the fact that, although a chance score would have been 8, the mean score was approximately 13.5 even before any interaction. In other words, the students could predict one another's responses with considerable accuracy even before they interacted! This matter will be discussed in greater detail later.

In another study, Bieri (1955a) investigated the relationship between the behavioral traits used in describing persons (cognitive complexity), and the operational constructs of *predictive accuracy* and "*assimilative*" *projection*. *Predictive accuracy* was composed of "accurate" projection (wherein "A" correctly predicts a response for "B" which is the same as "A's" self prediction), and accurately perceived differences between forecaster and forecastee. "*Assimilative*" *projection* was composed of "accurate" projection and "inaccurate" projection.

The results in Table 4 seem to

TABLE 4

CORRELATIONS BETWEEN COGNITIVE COMPLEXITY AND VARIOUS PERCEPTUAL SCORES (BIERI, 1955)

Predictive Behavior	Cognitive Complexity
	r
Predictive accuracy	.29
"Assimilative" projection	−.32*
"Accurate" projection	.02
Accurate perceived differences	.35*
"Inaccurate" projection	−.40*
Actual similarity	.20

* Significant at .05 level.

show that lack of cognitive complexity in the perception of others leads to the perception of others as similar in response habit to the forecaster, often with a resulting loss in accuracy of perception. The study is somewhat weakened by the fact that several of the correlations are predetermined, and nonindependent from the other correlations. Thus, given that the correlation involving *accurate perceived differences* is positive, it is most probable that the correlation containing "*inaccurate*" *projection* is negative. The reason for this may be readily seen. Considering all of the responses where the forecaster and forecastee differ in their self-responses, the following truism emerges: if the forecaster accurately perceives these differences, then he does not inaccurately perceive the forecaster as similar to himself. In other words, he does not perceive inaccurately. Hence, anyone with a high accurate perceived differences score must consequently have a low "inaccurate" projection score. If both variables are themselves correlated with the same variable (cognitive complexity of perception), then a positive correlation for *accurate perceived differences* (.35) makes it probable that there will be a negative correlation for "inaccurate" projection (−.40). In the foregoing table, therefore, few correlations may be selected without determining the others to some degree.

In another study, Bieri, Blacharsky, and Reid (1955) used the Incomplete Sentences Blank (ISB), and the Manifest Anxiety Scale (MAS), as indicators of adjustment. They predicted a negative relationship between degree of maladjustment and accuracy of predicted behavior, and a positive relationship between the degree of maladjustment and the tendency towards "assimilative" projection. The *S*s were 33 college men and 7 college women. The results were opposite to those predicted. The better adjusted individuals used "assimilative" projection more than the more poorly adjusted, while the poorly adjusted were significantly more accurate in perceiving differences between themselves and others. The better adjusted tended to perceive others like themselves ($r = .71$) and they were largely accurate in this perception ($r = .64$). The more maladjusted persons tended to perceive others as different from themselves with high accuracy. *Thus, one is faced with two different subgroups having good predictive ability with regard to other persons, one group being called "projectors," and the other not.* What more concise proof of the inadequacy of the definition of projection? The measure of projection is again a function of the homogeneity of the group. In a fairly homogeneous college group there is a high degree of similarity between responses to various questionnaires, and accordingly a high degree of "projection" and "accuracy" for those able to discern group homogeneity. Deviates in terms of maladjustment are readily aware of the differences between themselves and the group, and can accordingly predict others' responses accurately. Such persons, however,

rarely "project," because they are quite different from the majority of the group. "Projection," in this sense, is an artifact, since for both adjusted and maladjusted persons the same degree of predictive efficiency receives different labels with different connotations.

The evidence from experiments utilizing "attributive" projection, therefore, would seem to support the following conclusions:

(*a*) The use of discrepancy scores has resulted in many psychological findings which are statistical artifacts.

(*b*) "Attributive" projection may be related to ego-involvement. Whether projection occurs or not would seem to be dependent upon the relation of such behavior to the self concept.

(*c*) "Attributive" projection may result from such diverse phenomena as correctly perceiving similarity between another and oneself, or incorrectly perceiving such similarity. It may stem either from a lack of information, often referred to as "naive" projection, or from an adequate supply of information (accurately perceived similarity). In short, the term connotes such a varied number of meanings that it possesses little explanatory significance without reference to the composition of the group of perceivers, their homogeneity, and ego-involvement in the task. Also to be considered is the variability of the "personalities" to be judged as well as the nature of the information available about them. The measurement of "attributive" projection is more likely, therefore, to reveal the cognitive-response habits of a judge than information about the dynamics of his personality.

"*Autistic*" *Projection*

One of the earliest recognized distortions in perception has been that which stems from the manifest needs of the subject. In 1904, Külpe found that "actual sensory qualities of the stimulus which were not relevant to the task-set were to all intents and purposes not seen" (Helson, 1953, p. 22). Helson also gives the example of the classical "complication" experiment where the *S* watches a moving pointer and reports its location at the moment a bell is sounded. The stimulus for which the *S* is "set" is perceived prior to the incidental one. "Thus if the bell sounds at 20 objectively, it is seen at scale division 10 when the pointer is attended to; if the sound is attended to, the pointer is seen at 30. This phenomenon is perceptual and is not a matter of judgment" (Helson, 1953, p. 22).

Murray (1933) in some early work preceding the development of the TAT, gave some photographs to a group of eleven-year-old girls at a house party which his daughter gave. The girls rated the pictures once after a normal pleasurable experience, and once after a game of "murder." There was a considerable increase in the degree of maliciousness attributed to the pictures after the second game. That even nonambiguous tasks may elicit projection of emotional states has been reported by Johnson (1937a, 1937b) who found that normal but euphoric persons tended to overestimate distances between two points, while normal but depressed subjects tended to underestimate these same distances.

Sanford's two studies (1936, 1937) on the effects of abstinence from food on imaginal processes were the precursors to the "New Look" in perception. He gave 10 children words to associate and ambiguous pictures to interpret, both before and after meals. The subjects gave significantly more food responses before a a meal than after one. McClelland

and Atkinson (1948) have also found increases in food responses as a function of periods of deprivation not exceeding 16 hours.

Levine, Chein, and Murphy (1942), using food deprivation periods of 3, 6, and 9 hours, found that a deprived group gave more food responses after 3 and 6 hours than at the start in response to chromatic ambiguous figures projected on a screen. After the 9th hour, however, they manifested a decrease. Similarly Brozek, Guetzkow, and Baldwin (1951) working with a group of men subjected to semistarvation for 24 weeks measured the perception of food responses through the use of direct questioning, by the Rosenzweig Picture-Frustration Test, the Kent-Rosanoff Free-Word Association Test, and the Rorschach. The number of food responses showed no significant change as a function of time. Apparently the projection of deprived need is not a simple phenomenon. Projection occurs only when the *S* believes there is some chance of immediate gratification. When it becomes apparent that gratification will be delayed there is no projection. Projection is once again seen to be subsidiary to the need-gratification expectation of the perceiver.

Rationalized" Projection

The data with regard to "rationalized" projection have supported this meaning of the concept in the few experiments reported. In a study at the University of Vienna, Frenkel-Brunswik (1939) had four judges rate the conduct of some 40 students as well as write a description of their personalities. The students wrote an autobiography concerning their conduct at the University, the principles guiding their conduct, and the changes they felt should be made. She found that many students felt that the environment should compensate for their personal shortcomings. "In this instance the subject seems not to realize his own defect, but rather to project it on the environment" (1939, p. 418). For example, the rho correlation between the rating of lack of scientific ability by the judges and comments by the students of a need for pedagogical changes was .60. A lack of discipline on the part of the students was rationalized by their demand for more regimentation (rho = .62). Moreover, the overambitious persons often checked "I always do what I am ordered to do," and the aggressive ones asserted "I do not let myself be intimidated."

In a unique and imaginative experiment, Posner (1940) gave a group of eight-year-old children two toys to play with, one preferred, and the other nonpreferred. They were then asked to give one to a friend to play with, after which each child was asked which toy he thought the friend would have given away. A selfish judgment (friend also would have given away the nonpreferred toy) was considered to indicate projection. The control situation, created to avoid guilt feelings, and involving another matched group, did not require the child to give away one of the toys. Under these circumstances there was a much smaller degree of projection of selfishness.

Bellak (1944) found that the aggressive word content increased in TAT stories in the last five cards, when subjects were severely criticized, after receiving the first five cards without comment. Although Bellak interpreted this as "true" (Freudian) projection, he reported that when criticized for the poor quality of their stories, the subjects *admitted their inability* but offered excuses. They blamed the "ambiguity" of the pictures and the "inadequacy" of the instructions for their

poor performance. Apparently, the subjects were producing "rationalized" projection rather than "classical" projection. Their stories manifested considerable hostility of which they were aware, but which they could not express because of the authoritiative position of the examiner (Murstein, 1957b).

Miscellaneous Studies of Projection

Several experiments could not be categorized either because of a lack of information concerning the concept of projection or because the aim of the experiment was to arrive at some refinement of categories (Cattell & Wenig, 1952). Holt (1951) asked 10 judges to rate each of 10 college students on 36 personality variables, including "projection" on a 6-point scale. "Self-insight" was the only concept not rated as such, but was obtained as the sum of the squared differences between the ratings of the judges and the subjects. The correlation between "projection" and "self-insight" was .50, which, because of the paucity of subjects, did not quite reach significance at the .05 level. The correlation of "projection" with the six most attractive traits was .54; with the six most unfavorable traits .04. "Intelligence" was found to correlate .48 with "projection." These rather unusual findings are difficult to interpret because one does not know what conception of projection each judge held. Another difficulty is the lack of information as to the relative number of favorable traits compared to unfavorable ones. If one hypothesizes that the judges were using an attributive concept of projection without reference to the self concept of the *S*, then the results are somewhat more understandable. In a fairly homogeneous select group (Harvard collegians), the smarter, perhaps better adjusted students, would tend to refer to others as similar to themselves (probably quite correctly), and to be perceived by the judges as "projectors," since they would have attributed characteristics to others that they themselves possessed. Once again, in this eventuality, one would deal with persons classified as "projectors" solely through the small amount of group variance resulting from group homogeneity.

The Blacky Test was used by Cohen (1956) on a college group in which *S*s were rated on the dimentions of projection, regression, reaction-formation, and avoidance. The *S*s worked together on a task and evaluated each other on interpersonal ability. Cohen found that "projectors" in a group showed more negativeness and hostility in interpersonal relationships towards each other than when in other groups. When paired with "nonprojectors," however, "projectors" were not more threat-oriented than other dissimilar pairs. The hierarchy of defense responses with regard to perceived negative interaction was (*a*) projection, (*b*) regression, (*c*) reaction formation, and (*d*) avoidance.

These interesting results indicate that projection, as measured by the Blacky Test, seemed to give an indication of noncommunication resulting from the perception of extreme threat to the self. It would have been helpful if the basis for deciding who was, and who was not, a projector had been discussed in greater detail.

Cattell has concerned himself with the various concepts of projection (1944, 1951). An ambitious attempt to factor-analyze some of the different conceptions of projection was made by Cattell and Wenig (1952). The authors used the term "misperception" because they felt that they were dealing with measurement to a discrepancy between individual reac-

tions to the TAT-like pictures, which they used, and some superindividual standard of reaction. They believed that "misperception" is affected by three considerations: (*a*) abilities, (*b*) experiences (information, skills), and (*c*) the dynamic needs of the subject. Their hypothesis was that the nature and magnitude of misperception effects generally are to be accounted for by eight misperception factors within the *S*s. These were: (*a*) cognitive (intelligence), (*b*) cognitive (information concerning the field), (*c*) cognitive (information concerning the principal person), (*d*) consciously accessible dynamic traits (autism), (*e*) press-compatibility, and the defense mechanisms: (*f*) projection, (*g*) rationalization, and (*h*) phantasy.

A factor analysis of the stories selected (8 choices were possible, one for each hypothesized factor), and other "marker" variables did not seem to bear out the hypothesis. Various factors were teased out of the data and given labels of "phantasy," "naïveté," "autism," "rationalization," and "true projection." Still, it is evident from the examination of loadings on each factor that several kinds of "misperception" are represented in each factor. On the "naïveté" factor, the "naïveté" measures range from .77 to .48. Nevertheless, a "press" loading of approximately .40, as well as an "autism" loading of equal magnitude, was present. The "autism" factor contained substantial loadings for "rationalization" as well as for "autism." The "true" projection factor was loaded with "press" projection, "phantasy," and "naïveté."

Cattell and Wenig seemed mostly concerned with the high "press" loading on the "true" projection factor. They stated that a single underlying mechanism might be involved both in projecting an unconscious drive and in reconciling external facts with conscious internal moods, namely, a need for self consistency. "If so, the underlying process would be better labeled as 'self-saving'. . . . Secondly, and far more important, the misperception is not due to a single process, projection, but to several dynamic misperception processes, which superimposed, frequently act in different directions" (1952, p. 807).

While Cattell and Wenig's conclusions were based on an extensive analysis of the data, there appears to be a serious difficulty with regard to their operational conception of what constituted the various kinds of "misperception" or "projection" as shown by these examples of what they regarded as different kinds of projection: the woman in the picture *is dominating the man because he wants her to* . . . he enjoys being dominated (autistic misperception); the man kneeling at the boy's side *is dominating the boy because the boy is a very submissive person which necessitates that he be led* (press-compatibility misperception); the older man is the boy's father who *dominates the boy for his own good* (rationalization).

One wonders whether any basic differences exist between these three themes. All of them seem to embody a "rationalized" type of projection. The difficulty in detecting clearly independent kinds of projection may very well be traced to the nonindependence of the choices presented to the *S*s. It would be of interest to repeat this study using valid operations for distinguishing the different types of projection.

Lastly, Jenkin (1956) had *S*s look at pictures of varying ambiguity projected on a screen. "Projection" was said to occur when the subject seemed sure of the objective reality of his interpretation. "Rationalization" occurred when the subject made

his interpretation in a tentative manner. Correlations of this version of projection with various Rorschach and Rosenzweig P-F determinants were reported. Projection was significantly correlated with the Rorschach's *M*% (.38), Rosenzweig's E% (.36), and E-D% (.69). Negative correlations were found for Rosenzweig's M% (−.35), and N-P% (−.54). "Rationalization" was positively correlated with Rosenzweig's M% (.42), and negatively correlated with his E% (−.38) and the Rorschach's *W*% (−.49), and *M*% (−.43). Obviously, "projection" here meant a confident assertive method of reporting perceptions while "rationalization" signified a Casper Milquetoast-like approach.

The correlation between projection and projection/rationalization was reported as .94, while that of rationalization with projection/rationalization was found to be −.40. Both are artifacts containing spurious common elements. In the first case, the spurious element in the numerator (projection) assured a positive correlation, while in the latter case, the common element in the denominator (rationalization), provided an artifactual negative correlation.

The "Operational" Dilemma

The adherents of the psychoanalytic school of personality measurement have been strongly attacked for their failure to define operationally such concepts as "projection" and "repression." These criticisms, justified to a large extent, have led the more recent psychodynamically-oriented researchers to embrace more operational definitions of such variables as "projection," "empathy," "reality," and "insight." Unfortunately, the result of zealous operationism has been a neglect of the original psychological meaning of these variables. Instead, we are faced with a string of operational phrases of the "A" rates himself ("A") as he thinks "B" would rate him—variety. When one has plodded through the adding and subtracting of the 8 operations necessary to arrive at a score of "Refined Empathy" or the 12 operations used in correlating "Refined Empathy" with "Projection" one may be sorely tempted to return to the "good old days" of literary definition. *What is needed is an operational definition which does not depart from the accepted psychological meaning of a concept.* It is clear that much of the research reviewed in this paper has failed to fufill this need. Instead, judges have been asked to predict, according to Gage and Cronbach (1955), such diverse operations as:

> (*a*) how *persons in general* will behave;
> (*b*) how *a particular category of persons* deviates from the behavior of persons in general;
> (*c*) how *a particular group* deviates from the typical behavior of the particular category it belongs to;
> (*d*) how *an individual* deviates from the typical behavior of the particular group he belongs to;
> (*e*) how *an individual on a particular occasion* will deviate from his typical behavior (Gage and Cronbach, 1955, p. 413).

One is faced here with many abilities whose relationship to personality patterns is rather slight. Cronbach (1955) mentions several of these abilities including "differential elevation," measuring the forecaster's ability to judge deviation of the individual's elevation from the average; "stereotype accuracy," measuring the judge's accuracy in predicting the "generalized other"; "differential accuracy," reflecting the judge's ability to predict differences between subjects on any item. The shrewd judge is often one whose predictive variance (σ_y) does not exceed the self predictive variance of the person being judged (σ_x). "Accuracy is improved

as σ_y approaches $r_{xy}\sigma_x$. That is to say, the variation in predictions should never exceed the variation in true responses and should ordinarily be much smaller" (Cronbach, 1955, p. 181).

Crow (1954) found predictions to be accurate for the generalized "other" person, when the group for whom predictions were made was (*a*) homogeneous and (*b*) judges exhibited response sets which were general over "others." In another study, Crow and Hammond (1957) found that the response set of medical students (perceptual stereotypy of patients' responses to a personality scale) was more consistent over time than differential accuracy (the ability of *S* to predict differences between other people on any item).

Eight keys to interpersonal perception are listed by Gage, Leavitt and Stone (1956) as affecting accuracy. The keys are hypothetical protocols derived from various constructs and compared with the forecaster's prediction for *S* and *S*'s self-response. The keys found to be important in determining responses were:

A Priori Keys:

1. *Acquiescent Tendencies on the Part of the Judge and S.* If both are highly acquiescent, the former in predicting, and the latter in answering personality inventories, accuracy of interpersonal preception will be high.

2. *Favorability of the Judges' Predictions and of the S's Self.* If the items possess high social desirability and if the *S* wishes to appear in a favorable light, there will again be high accuracy of perception by the judges.

3. *Adjustment Tendencies Derived from Adjustment Value of Responses.* These are closely allied to the "favorability" tendencies.

Keys Obtained by Varying Instruction to Judges:

4. *Judge's Self-Description.* Judges may be accurate because of high similarity to the subject and their assuming similarity in their predictions, or they may be dissimilar, and assume little similarity and still be accurate.

5. *Stereotypy.* By predicting the typical member, a stereotypy key is obtained. Accuracy ensues when the judge follows his stereotype when the subject is closely similar to it.

6. *A manifest stimulus value key* may be derived from the modal judge's descriptions of the subject. This key relates to the impression *S* makes on those judging him. When combined with *S*'s self perception, it yields an "insight" or "frankness" score.

Keys Based on Central Tendencies of Predictions or Self Descriptions

7. *Modal Prediction keys* may stem from (*a*) the average prediction of *n* judges for a single subject, and (*b*) the average prediction by a single judge for many subjects. The former key will result in high accuracy when judges make highly typical predictions for highly predictable others. High accuracy also will occur when a judge makes an atypical correct prediction for the way *S* will answer an item, while the majority of judges make a stereotyped but incorrect choice. The latter key (*b*) gives a measure of a judge's "implicit stereotype."

8. *A modal self-description* of the subjects may be utilized against a single subject's self-responses and a single judge's prediction. Accuracy then depends upon the judge's stereotype of the group as well as the individual's deviation from the stereotype of the group.

It is apparent that these various "keys" do not reflect very much

about defense mechanisms. Instead, they seem to consist of cognitive response habits in which a fastidiousness for operationism has resulted in a growing aridness of conceptual meaning. One may well speculate as to the multitude of factors which would result from a factor analysis of the current operational definitions of "projection." It is therefore hardly remarkable that projection has been found to correlate both positively and negatively with such variables as "insight" and "empathy." The dilemma can be resolved by insisting that operationism be utilized jointly with clinical meaning rather than merely supplanting the latter.

Discussion

From birth, man views the world imperfectly, using his sensitive but far from perfect eye as a camera to bring him into communication with the outer world. While it is true that states correlated with cortical impulses are projected to the supposed location of the stimulation, such a concept of projection seems too broad to be of use. We propose to limit the definition of projection to perception or judgments having to do with the personality of the organism, thereby eschewing any physiological or cognitive components. A definition which is broad enough to cover the material reviewed in this article is as follows:

Projection: The manifestation of behavior by an individual which indicates some emotional value or need of the individual.

Such behavior may vary in the degree of defensiveness, depending upon the situational context and the personality of the perceiver. Thus, where little or no ego-threat is involved, a person may project his emotional values through his method of organizing and selecting his personal milieu and the objects inhabiting it. One might term such behavior as "life style" projection. The fact that these behaviors are public usually indicates that they involve little threat to the "self." "Autistic" projection illustrates the reaction to strong needs, but does not necessarily mean these needs are defensive ones. For one who eagerly awaits a visit from Aunt Agatha, who will be wearing a green coat, any middle-aged woman emerging from the train wearing a green coat may be momentarily misperceived as the expected relative. In "rationalized" projection, however, we deal with data in which, despite the fact that the *content* is readily accessible to consciousness, *the motivation is distorted* so as to deceive with regard to the real intent of the act. The peak of distortion, "classical" projection, occurs when *even the content must be denied* because of the extent of the threat,and the individual sees the unwanted behavior as stemming wholly from outside sources.

It will be noted that the concept of "attributive" projection has been omitted from this discussion. The reasons for so doing are that the operation of saying another would do as the *S* would do, need not involve any emotional values. It may stem instead from (*a*) a shrewd knowledge of the other person's behavioral tendencies (perceiving correctly actual similarities between the "other" and oneself), (*b*) dull intellect (perceiving another as similar when he is not), (*c*) lack of information ("naïve" projection), or (*d*) the differing needs of the person, which are either relatively nondefensive ("autistic" projection), or strongly defensive ("rationalized" or "classical" projection).

In general, we have been critical of the "operational" school of researchers who have attempted to measure projection on the basis of a

forecaster's response regarding himself and others through the use of personality inventories and questionnaires. The correlations accruing from these studies have usually been spurious in containing identical operations in the variables correlated. Moreover, the number of correlations have often exceeded the available degrees of freedom.

In addition to being parsimonious, the operation measuring projection should be closely allied to the *personality construct* being investigated so that one is not confronted with the phenomenon of coldly objective, uninvolved persons mysteriously emerging as "projectors." Projection should not be a function of cognition, but of emotional involvement.

In achieving the goals set forth, the measuring instruments should embody the following considerations:

(*a*) The instrument should have a low internal consistency for any possible intermediary keys not related to the investigation.

(*b*) If the composition of the experimental group is in the hands of the investigator, the group should be made as heterogeneous or representative as possible (Crow, 1957a). This would also check against a lucky guess in which one person was generalized for a similar population, thus giving a spuriously high predictive efficiency score. The representativeness of a group might also prevent spuriously poor predictions when several individuals judge only one person who well may be an atypical member of his group.

(*c*) It is extremely unlikely that a standard stimulus (ambiguous picture) will elicit the same personality-meaningful reaction in every person. For example, if as a result of short-term food deprivation a large majority of persons see food or instrumental activity leading to food, we do not have a manifestation of projection. Rather, by the very fact that most persons react in the same manner, we may conclude that we deal with a biosocial reaction with little meaning for the understanding of the emotional life of a given individual.

(*d*) The relationship between the information possessed by *S* and the kind of judgment asked of him should be investigated more thoroughly. Presumably the more "internal" the perception of others, the less relevant the information fed to the perceiver need be to induce inappropriate perception.

(*e*) If projection is to be a phenomenon involving emotional values with regard to the self, it is the task of the experimenter to indicate in what manner the judgments of the perceiver are indicative of these values rather than of relatively extraneous cognitive-response habits.

In closing, one must allow much credit to the earlier workers involved with the concept of projection. Their bold sallies into unknown and tricky terrain have stimulated more extensive work in this area. Many subtle artifacts have been painstakingly removed from the material supposedly concerned with "projection," "empathy," "similarity," "reality," and "insight." One might guess that we are about ready for a new and more sophisticated set of personality constructs. Our position is like that of the harried Christmas shopper who knows quite clearly what he does not want, but is not quite sure where to go from that point.

REFERENCES

ALFERT, ELIZABETH. Two components of assumed similarity. *J. abnorm. soc. Psychol.*, 1958, **56**, 135–138.

ALLPORT, G. S. *Personality: A psychological interpretation.* New York: Holt, 1939.

ATKINSON, J. W., & MCCLELLAND, D. C. The

projective expression of needs. II. The effect of different intensities of the hunger drive on thematic apperception. *J. exp. Psychol.*, 1948, **38**, 643–658.

BALDWIN, A. L. *Behavior and development in childhood.* New York: Dryden, 1955.

BELL, J. E. *Projective techniques.* New York: Longmans, Green, 1948.

BELLAK, L. The concept of projection. *Psychiatry*, 1944, **7**, 353–370.

BELLAK, L. On the problems of the concept of projection: A theory of apperceptive distortion. In L. Abt and L. Bellak (Eds.) *Projective psychology*, New York: Knopf, 1950. Pp. 7–32.

BELLAK, L. *The Thematic Apperception Test and The Children's Apperception Test in clinical use.* New York: Grune & Stratton, 1954.

BELLAK, L. Freud and projective techniques. *J. proj. Tech.*, 1956, **20**, 5–14.

BENDER, I. E., & HASTORF, A. H. The perception of persons: Forecasting another person's responses on three personality scales. *J. abnorm. soc. Psychol.*, 1950, **45**, 556–561.

BENDER, I. E., & HASTORF, A. H. On measuring generalized empathic ability (social sensitivity). *J. abnorm. soc. Psychol.*, 1953, **48**, 503–506.

BIERI, J. Changes in interpersonal perceptions following social interaction. *J. abnorm. soc. Psychol.*, 1953, **48**, 61–66.

BIERI, J. Cognitive complexity-simplicity and predictive behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 263–268.

BIERI, J., BLACHARSKY, E., & REID, J. W. Predictive behavior and personal adjustment. *J. consult. Psychol.*, 1955, **19**, 351–355.

BROZEK, J., GUETZKOW, H., & BALDWIN, M. W. A quantitative study of perception and association in experimental semi-starvation. *J. Pers.*, 1951, **19**, 245–264.

CALVIN, A. D., & HOLTZMAN, W. H. Adjustment and the discrepancy between self concept and the inferred self. *J. consult. Psychol.*, 1953, **17**, 39–44.

CAMERON, N. *The psychology of behavior disorders.* New York: Houghton Mifflin, 1947.

CAMERON, N. Perceptual organization and behavior pathology. In R. R. Blake and G. V. Ransey (Eds.), *Perception: An approach to personality.* New York: Ronald, 1951. Pp. 283–306.

CATTELL, R. B. Projection and the design of projective tests of personality. *Charact. & Pers.*, 1944, **12**, 177–194.

CATTELL, R. B. Principles of design in "projective" or misperception tests of personality. In H. H. Anderson and Gladys L. Anderson (Eds.), *An introduction to projective techniques.* New York: Prentice-Hall, 1951 Pp. 55–98.

CATTELL, R. B., & WENIG, P. W. Dynamic and cognitive factors controlling misperception. *J. abnorm. soc. Psychol.*, 1952, **47**, 797–809.

COHEN, A. R. Experimental effects of ego-defense preference on interpersonal relations. *J. abnorm. soc. Psychol.*, 1956, **52**, 19–27.

COWDEN, R. C. Empathy or projection? *J. clin. Psychol.*, 1955, **11**, 188–190.

CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.

CROW, W. J. A methodological study of social perceptiveness. Unpublished doctoral dissertation, Univer. of Colorado, 1954.

CROW, W. J. The need for representative design in studies of interpersonal perception. *J. consult. Psychol.*, 1957, **21**, 322–325. (a)

CROW, W. J. The effect of training upon accuracy and variability in interpersonal perception. *J. abnorm. soc. Psychol.*, 1957, **55**, 355–359. (b)

CROW, W. J., & HAMMOND, K. R. The generality of accuracy and response sets in interpersonal perception. *J. abnorm. soc. Psychol.*, 1957, **54**, 384–390.

DYMOND, ROSALIND F. Personality and empathy. *J. consult. Psychol.*, 1950, **14**, 343–350.

FABIAN, W. A. An investigation of the relationship between measures of insight and measures of projection and distortion in ratings. Unpublished doctoral dissertation, Univer. of Buffalo, 1954.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

FIEDLER, F. E. A method of objective quantification of certain counter-transference attitudes. *J. clin. Psychol.*, 1951, **7**, 101–107.

FIEDLER, F. E. The psychological distance dimension in interpersonal relations. *J. Pers.*, 1953, **22**, 142–150.

FIEDLER, F. E., BLAISDELL, F. J., & WARRINGTON, W. G. Unconscious attitudes and the dynamics of sociometric choice in a social group. *J. abnorm. soc. Psychol.*, 1952, **47**, 790–796.

FRENKEL-BRUNSWIK, ELSE. Mechanisms of self-deception. *J. soc. Psychol.*, 1939, **10**, 408–420.

FREUD, S. *Totem and taboo and other works.* London: Hogarth, 1955.

FRIEDMAN, I. Phenomenal, ideal, and projected conceptions of self. *J. abnorm. soc. Psychol.*, 1955, **51**, 611–615.

GAGE, N. L., & CRONBACH, L. J. Conceptual

and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411–422.

GAGE, N. J., LEAVITT, G. S., & STONE, G. C. The intermediary key in the analysis of interpersonal perception. *Psychol. Bull.*, 1956, **53**, 258–266.

GOLDINGS, H. J. On the avowal and projection of happiness. *J. Pers.*, 1954, **23**, 30–47.

GOSS, A. E., & BROWNELL, MARJORIE H. Stimulus-response concepts and principles applied to projective test behavior. *J. Pers.*, 1957, **25**, 505–523.

HALPERN, N. M. Empathy, similarity, and self-satisfaction. *J. consult. Psychol.*, 1955, **19**, 449–452.

HASTORF, A. H., & BENDER, I. E. A caution respecting the measurement of empathic ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 574–576.

HASTORF, A. H., BENDER, I. E., & WEINTRAUB, D. J. The influence of response patterns on the "Refined Empathy" Score. *J. abnorm. soc. Psychol.*, 1955, **51**, 341–343.

HEALY, W., BRONNER, AUGUSTA, P., & BOWERS, A. M. *The structure and meaning of psychoanalysis.* New York: Knopf, 1930.

HELSON, H. Perception and personality—a critique of recent experimental literature. *Air. Univ. Sch. Aviat. Med.*, 1953, Proj., No. 21-0202-0007, Rep. No. 1.

HOFFMAN, E. P. Projection and Ich-Entevicklung. *Int. Z. Psychoanal.*, 1935, **21**, 342–373.

HOLT, R. R. The accuracy of self-evaluations: Its measurement and some of its personological correlates. *J. consult. Psychol.*, 1951, **15**, 95–101.

HOOP, J. H., VAN DER. "Ueber du Projektion und ihre Inhalte." *Int. Z. Psychoan.*, 1924, **10**, 276–288.

HORNEY, KAREN. *New ways in psychoanalysis.* New York: Norton, 1939.

JANET, P. Charactères de l'hallucination du persecuté. In A. Michotte, (Ed.) *Miscellanea Psychologica.* Louvain: Inst. sup. de Phil., 1947.

JELGERSMA, G. Projection. *Int. J. Psychoan.*, 1926, **7**, 353–358.

JENKIN, N. S. Some relationships between projective test behavior and perception. *J. clin. Psychol.*, 1956, **11**, 278–281.

JOHNSON, WINIFRED B. Euphoric and depressed moods in normal subjects: Part I. *Charact. & Pers.*, 1937, **6**, 79–98. (a)

JOHNSON, WINIFRED B. Euphoric and depressed moods in normal subjects: Part II. *Charact. & Pers.*, 1937, **6**, 188–212. (b)

KAUFMAN, M. R. Projection, heterosexual and homosexual. *Psychoanal. Quart.*, 1934, **3**, 134–136.

KNIGHT, R. P. Introjection, projection, and identification. *Psychoanal. Quart.*, 1940, **9**, 334–341.

LECKY, P. *Self-Consistency.* New York: Island, 1951.

LEVENTHAL, J. Cognitive processes and interpersonal predictions. *J. abnorm. soc. Psychol.*, 1957, **55**, 176–180.

LEVINE, R., CHEIN, I., & MURPHY, G. The relation of the intensity of a need to the amount of perceptual distortion: A preliminary report. *J. Psychol.*, 1942, **13**, 283–293.

LUNDY, R. M., Assimilative projection and accuracy of prediction in interpersonal perceptions. *J. abnorm. soc. Psychol.*, 1956, **52**, 33–39.

LUNDY, R. M., & BERKOWITZ, L. Cognitive complexity and assimilative projection in attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 34–37.

MCCLELLAND, D. C., & ATKINSON, J. W. The projective expression of needs: I. The effect of different intensities of the hunger drive on perception. *J. Psychol.*, 1948, **25**, 205–222.

MINTZ, ELIZABETH. An example of assimilative projection. *J. abnorm. soc. Psychol.*, 1956, **52**, 279–280.

MUHL, ANITA M. Identification-projection-domination. *Med. Wom. J.*, 1943, **50**, 219–228.

MUNN, N. L. *Psychology.* New York: Houghton Mifflin, 1946.

MURPHY, G. *Personality.* New York: Harper, 1947.

MURRAY, H. A. The effect of fear upon estimates of the maliciousness of other personalities. *J. soc. Psychol.*, 1933, **4**, 310–329.

MURRAY, H. A. *Explorations in personality.* New York: Oxford ,1938.

MURRAY, H. A. Foreword in H H. Anderson and Gladys L. Anderson (Eds.) *An introduction to projective techniques.* New York: Prentice-Hall, 1951.

MURSTEIN, B. I. The projection of hostility on the Rorschach and as a result of ego-threat. *J. proj. Tech.*, 1956, **20**, 418–428.

MURSTEIN, B. I. Some comments on the measurement of projection and empathy. *J. consult. Psychol.*, 1957, **21**, 81–82. (a)

MURSTEIN, B. I. Studies in projection: A critique. *J. proj. Tech.*, 1957, **21**, 129–136. (b)

NORMAN, R. D., & AINSWORTH, PATRICIA. The relationships among projection, empathy, reality, and adjustment, operationally defined. *J. consult. Psychol.*, 1954, **18**, 53–58.

NORMAN, R. D., & LEIDING, W. C. The rela-

tionship between measures of individual and mass empathy. *J. consult. Psychol.*, 1956, **20**, 79–82.

PAGE, H. A., & MARKOWITZ, GLORIA. The relationship of defensiveness to rating scale bias. *J. Psychol.*, 1956, **41**, 431–435.

PIAGET, J. *La représentation du monde chez l'enfant.* Paris: Presse Univer. de France, 1926.

POSNER, B. A. Selfishness, guilt feeling and social distance. Unpublished master's thesis, Univer. of Iowa, 1940.

Psychiatric Dictionary. New York: Oxford, 1940.

RAPAPORT, D. Principles underlying projective techniques. *Charact. & Pers.*, 1942, **10**, 213–219.

RAPAPORT, D. Projective techniques and the theory of thinking. *J. proj. Tech.*, 1952, **16**, 269–275.

RAPAPORT, D., SCHAFER, R., & GILL, M. *Manual of diagnostic psychological testing.* II. Diagnostic testing of personality and ideational content. (2 Vols.) Chicago: Year Book, 1945.

ROKEACH, M. Studies in beauty: II. Some determiners of the perception of beauty in women. *J. soc. Psychol.*, 1945, **22**, 155–169.

SANFORD, R. N. The effects of abstinence from food upon imaginal processes: A preliminary experiment. *J. Psychol.*, 1936, **2**, 129–136.

SANFORD, R. N. The effects of abstinence from food upon imaginal processes: A further experiment. *J. Psychol.*, 1937, **3**, 143–159.

SCHACHTEL, E. G. Projection and its relation to character attitudes and creativity in the kinesthetic responses. *Psychiatry*, 1950, **13**, 69–100.

SCHAFER, R. *Psychoanalytic interpretation in Rorschach testing.* New York: Grune & Stratton, 1954.

SEARS, R. R. Experimental studies of projection: I. Attribution of traits. *J. soc. Psychol.*, 1936, **7**, 151–163.

SEARS, R. R. Experimental studies of projection: II. Ideas of reference. *J. soc. Psychol.*, 1937, **8**, 389–400.

SEARS, R. R. Survey of objective studies of psychoanalytic concepts. *Soc. Sci. Res. Counc. Bull.*, 1943, No. 51.

SHAFFER, L. F. *The psychology of adjustment.* New York: Houghton Mifflin, 1945.

SUCHMAN, J. R. Social sensitivity in the small task-oriented group. *J. abnorm. soc. Psychol.*, 1956, **52**, 75–83.

SYMONDS, P. *Dynamic psychology.* New York: Appleton-Century-Crofts, 1949.

THOMSEN, A. Psychological projection and the election: A simple class experiment. *J. Psychol.*, 1941, **11**, 115–117.

VAN LENNEP, D. J. The Four Picture Test. In H. H. Anderson and Gladys L. Anderson (Eds.), *An introduction to projective techniques.* New York: Prentice-Hall, 1951. Pp. 149–180.

VAN LENNEP, D. J. Projection and personality. In H. P. David and H. von Bracken (Eds.), *Perspective in personality theory.* New York: Basic Books, 1957. Pp. 259–277.

WALLEN, R. Individual estimates of group attitudes. *Psychol. Bull.*, 1941, **38**, 539–540.

WARREN, H. C. *Dictionary of psychology.* Boston: Houghton Mifflin, 1934.

WEINGARTEN, ERICA M. A study of selective perception in clinical judgment. *J. Pers.*, 1949, **17**, 369–396.

WEISS, E. Projection, extrojection and objection. *Psychoanal. Quart.*, 1947, **16**, 357–377.

WITTICH, J. J. The generality of the prediction of self-reports. *J. consult. Psychol.*, 1956, **19**, 445–448.

ZILBOORG, G. *The medical man and the witch during the Renaissance.* Baltimore: Johns Hopkins, 1935.

ZIMMER, H. The roles of conflict and internalized demands in projection. *J. abnorm. soc. Psychol.*, 1955, **50**, 188–192.

ZUCKER, K. Experimental investigation of correlates of projection. Unpublished master's thesis, Western Reserve Univer., 1952.

Received September 22, 1958.

PSYCHOLOGICAL BULLETIN
VOL. 56, NO. 5, 1959

RECENT ADVANCES IN THE CHEMISTRY OF PSYCHOTIC DISORDERS

LEONARD S. RUBIN
Eastern Pennsylvania Psychiatric Institute

An astonishing number of advances have been made recently by pharmacologists, biochemists, and neurophysiologists into possible causative chemical factors of schizophrenia and other psychoses. As a result of the reported efficacy of the "tranquilizers" and "psychic energizers" in the management of some psychotic patients, many plausible fruitful hypotheses have been generated and very often verified. As these advances continue to be made, the conviction grows among workers in the area that the organic (chemical) changes recently discovered in schizophrenia are relevant to its etiology.

The purpose of this review is to make available an introductory overview of several of the outstanding theoretical formulations of the biochemistry of psychotic behavior for the experimental as well as the clinical psychologist. These theories should also prove useful in providing some rational basis for several chemotherapies currently employed with psychiatric patients. However, prior to any such consideration it is of interest to review the results of some broad biochemical investigations which in the absence of specific hypotheses nevertheless uncovered interesting correlates which await integration into an as yet undefined unified biochemical theory of schizophrenia.

Some Physiologic and Biochemical Correlates of Psychotic Behavior

Any attempt to but briefly describe all or even a significant number of the studies that have shown correlation between physiological or biochemical factors and psychotic behavior would necessitate a separate review or even a volume. Instead, a limited number of studies will be cited that are representative of the broad spectrum of research areas that have been pursued intensively. Some workers have focused their attention on the possibility that schizophrenia may be due to a cerebral toxin which is the by-product of a metabolic deficit. Thus, it has been reported (Fischer, 1953) that the blood of schizophrenics was toxic to tadpoles while Federhoff and Hoffer (1956) reported that schizophrenic blood was toxic to fibroblasts nurtured in a tissue culture. However, this toxicity does not appear to be due to a toxic chemical specifically found in schizophrenic blood because the latter workers also found that the blood of patients undergoing surgery was equally toxic. This finding does suggest that some chemical, toxic to tadpoles and fibroblasts, is found in the bloodstream of people exposed to severe stress and is also found in the blood of schizophrenics. Reider (1957) employed Witt's spider-web technique to determine the effect of schizophrenic urine on the complex perceptuomotor organization required in web building and found significant abnormalities in the web pattern. Örström (1951) has focused his attention on phosphorus metabolism and has reported a diminished turnover of adenosine triphosphate and a higher content of phosphoglycolic acid in the erythrocytes of schizophrenics. Still another after-

math of metabolic impairment may be reflected in lowered concentrations of carbonic anydrase in the occipital cortex of schizophrenics and in early cases in the frontal cortex as reported by Ashby (1950). Richter (1957) has stated that the different enzyme systems of the liver seem to be impaired in schizophrenics and that different enzyme systems may be affected in varying degrees. However, he emphasizes that the enzymatic malfunction need not be the invariable concomitant of schizophrenia.

In addition to these findings of disturbed enzymatic reactivity or metabolic defect, there are physiological studies that have employed measurements of body temperature, heart rate, basal metabolic rate, and thyroid activity in normals and schizophrenics which demonstrate a greater variance of the measurements in schizophrenics than in healthy people. These results suggest that the schizophrenias represent a generalized disorder of steady-states. In summary, the few empirical findings reviewed up to this point demonstrate that schizophrenia may be characterized by some disordered metabolic response to stress, a defect in several important steps in phosphorus metabolism, a disturbance in the balanced distribution of carbonic anhydrase in the cerebral cortex, impaired enzymatic activity of the liver and finally, a disordered homeostasis which is reflected in the measurement of various steady-states.

"M-Substance"

An important biochemical theory of schizophrenia was formulated within this decade by Osmond and Smythies (1952), and elaborated by Hoffer, Osmond, and Smythies, (1954). The reader may be interested in the logical development of this theory. It was Hoffer (1957) who claimed that schizophrenia is a disease of the autonomic nervous system. Although vague as to the nature of the primary defect, whether it is a constitutional biochemical factor or psychogenic, the consequent events are each stated with warrantable assertability. A dominant feature of the autonomic disturbance is an increased parasympathetic activity or pronounced central cholinergic activity. This increased production of acetylcholine, by stimulating sympathetic ganglia, in turn produces an increase in the secretion of norepinephrine and epinephrine. Whereas in normal individuals epinephrine is metabolized by amine oxidase or sulfoesterase, in schizophrenia defective metabolism of epinephrine, by a phenolase rather than the amine oxidase or sulfoesterase, results in the production of quinone indoles (adrenochrome and adrenolutin) which interfere with cerebral metabolism. To prove this assertion that a cerebral toxin which is the by-product of some faulty metabolic process is the primary etiological factor in schizophrenia requires that a metabolite be isolated from schizophrenics which when introduced into healthy people should produce schizophrenia; or, the toxic metabolite should be found only in those people afflicted with schizophrenia; or, the toxin should be found in larger quantities; or, greater sensitivity to the metabolite by schizophrenics must be demonstrated. The next step in their analysis required the isolation of one metabolic fault out of the myriad metabolic processes that characterize life. They approached this problem by emphasizing the reported relationship between certain classes of chemical compounds and disordered behavior. Specifically, the "hallucinogens" had

the property of producing in healthy people affects judged similar to the symptoms of schizophrenia. It should be emphasized that these theorists did not equate schizophrenia with the state produced by LSD (lysergic acid diethylamide) or mescaline. They viewed the hallucinogens as capable of engendering a behavioral complex which could serve as a "model psychosis." The last step in the theoretical formulation required the identification of some endogenous metabolite which had a chemical structure similar to that of mescaline and the theorists pointed to adrenaline. However, as adrenaline is an essential neurohumor, it was proposed that derivatives of adrenaline produced as a result of faulty metabolism may serve as the "M-Substance," the mescaline-like toxin of schizophrenia. Axelrod (1957) was soon to provide a clue as to the manner in which the substance could be formed when he found that methylation of the phenolic groups of adrenaline took place in vivo. Resnick, Wolfe, Freeman, and Elmedilah (1958) reported that adrenaline is almost entirely detoxified by the process of O-methylation in human beings. These chemical findings tended to support the speculations of the early theorists that faulty metabolism of adrenaline, possibly excessive methylation, could produce the toxin in schizophrenics.

The quest for the exact description of "M-Substance" was given further impetus when it was claimed that "pink adrenaline" could produce psychotic-like reactions in healthy people when injected. "Pink adrenaline," it should be explained, is the resultant mixture of auto-oxidized products of adrenaline obtained when adrenaline is left exposed to light and air for some period of time. It was now suggested that one of the oxidized products such as adrenochrome or adrenolutine (trihydroxy N-methyl indole) or some other component of the oxidized mixture was the endogenous toxin, the "M-Substance." In the presence of many stabilizing substances in the blood, however, it is difficult to conceive of auto-oxidation of epinephrine in vivo. Experiments have been performed in an effort to identify the active component of the oxidized mixture which is responsible for the production of disturbed behavior. Obviously, it is to be expected from the foregoing that the injection of some toxin of the oxidation products of adrenaline should produce aberrant behavior in healthy people, or, the toxin should be found in schizophrenics and not in healthy people, or a significant quantitative difference should be found. Hoffer (1957) published observations on adrenolutine, an isomer of adrenochrome which is more stable. He tried to determine which persons had been given adrenolutine and which the placebo with equivocable results. Gastaldi (1957) attempted to ascertain the effect of the more highly reactive, unstable adrenochrome and reported the absence of behavioral abnormalities. In a recent study (Holland, Cohen, Goldenberg, Sha, & Leifer, 1958) equilibrium plasma concentrations of adrenaline and noradrenaline were measured and no significant difference between healthy people and schizophrenics with regard to the rate of utilization of circulating adrenaline and noradrenaline was found. Axelrod (1958) attempted to detect the presence of adrenochrome in schizophrenics and found that it was absent in schizophrenics as well as healthy people.

Hoffer (1957) deduced from his theory that chemical interference in the production of epinephrine from methylation of norepinephrine should

effectively reduce the amount of adrenochrome and thus lead to significant improvement in psychiatric patients. As nicotinic acid and nicotinamide are potent methyl group acceptors in vivo, he administered these compounds, and reported considerable success in the treatment of acute schizophrenics and little or no success with chronic schizophrenics. These results do not necessarily substantiate the theory.

The disappointing results of these experiments have been rationalized by the theorists in at least two ways. First, it has been argued that so unstable and biologically active a material as adrenochrome if introduced into the blood stream may not survive unchanged by the time it reaches the brain. Second, the point is made concerning the relevance of the blood-levels of the various endogenous metabolites to the activity of these toxins in the brain. For, if adrenaline and noradrenaline are neurohumors, then the substances required for their synthesis and destruction are also in the brain. Faulty metabolism of these catechol amines could produce toxins locally which would not be reflected in the blood because the toxins could not pass the blood-brain barrier.

The evidence in support of the prominent role ascribed to adrenochrome is based on some physiological observations on animals and indirect clinical evidence. Schwarz, Wakin, Bickford, and Lichtenheld (1956) reported that intraventricular injections of adrenochrome produced a drowsy state in monkeys with manifest heightened thresholds to painful stimuli. Walaszek, Smith, and Minz (1958) found that the serum of schizophrenics either abolished or reversed the systemic vasopressor response produced by topical application of adrenaline to the exposed cerebral cortex of the rabbit, whereas the serum of healthy people had no such effect. Indirect clinical evidence to support the role of adrenochrome in schizophrenia was provided by Lea (1955). He inferred that since adrenochrome is antihistaminic in action, if schizophrenics have abnormally excessive quantities in their circulation, then they should be more resistant to allergic conditions than healthy people. In his study of military schizophrenics, using head-injured soldiers as controls, he found a highly significant deficiency of allergic reactions among the schizophrenics.

In summary, the theory which postulates the existence of an "M-Substance," an endogenous metabolite of adrenaline produced by a deficiency in the process of methylation, as yet lacks the empirical verification for the criteria requisite for confirmation of the theory as established by the early theorists. Evidence is still forthcoming that the production of schizophrenic symptoms can be elicited in healthy people by a metabolite that is isolable from schizophrenics, or, that the metabolite is found solely in schizophrenics, or at least in significantly larger amounts in them as compared to healthy people.

Antimetabolites of Serotonin

While the previous biochemical theory of schizophrenia, it would seem, was established by an observation of the similarity in chemical structure between an hallucinogen, mescaline, and the neurohumors, adrenaline and noradrenaline, the second theory here to be considered was formulated after the observation was made that the most powerful hallucinogen, LSD, was a potent peripheral antagonist of serotonin (5-hydroxytryptamine). The role of

serotonin as a chemical mediator of neural activity in the brain has been described by Brodie and Shore (1957). The clue as to its important role came from its discovery in the brain. Like acetylcholine, the substance is present in nervous tissue in a precursor state and is active only in an unbound form. The amine is unevenly distributed in the brain, its concentration being highest in the brainstem, especially the hypothalamus, lowest in the cortex, and almost undetectable in the cerebellum. The high biological activity of this amine strongly suggests an important function in chemical mediation in subcortical centers. Additional support for a role of serotonin in neural transmission comes from consideration of the distribution of the enzyme monoamine oxidase, the enzyme that destroys serotonin, and 5-hydroxytryptophan decarboxylase, the enzyme that synthesizes serotonin, both of which are found in highest concentration in the hypothalamus.

Woolley and Shaw (1954) postulated that the hallucinogenic effect of LSD might be due to its interference with the action of serotonin centrally, in the brain. An excess of serotonin or a deficiency may lead to transmission dysfunction and concomitant behavioral disorder. According to Woolley and Shaw, the action of LSD, as well as several other alkaloids, can be ascribed to the indole moiety in the structures of LSD and serotonin. Consequently, compounds containing the indole ring may act as antimetabolites to serotonin. LSD is enough like serotonin in structure to be taken up by the serotonin receptors in the brain in lieu of serotonin. This suggests that mental aberrations are produced by an inadequate amount of serotonin at its site of action. However, these same workers (Shaw & Woolley, 1956) also showed in a later paper that LSD could act like serotonin in potentiating its effect on the blood pressure of anesthetized dogs. This then suggested the possibility that there may be an excessive concentration of serotonin in other forms of mental disease.

Several studies can be cited which tend to support the theory that interference in the activity of serotonin in the brain (possibly by an endogenous antimetabolite containing the indole nucleus) is related to psychotic behavior. In one study (Zeller, 1958) it was found that after the administration of large quantities of tryptophan to schizophrenics and healthy people, the schizophrenics excreted significantly less 5-hydroxyindolacetic acid than the normals. This acid is a break-down product of serotonin which in turn is a derivative of tryptophan.

The serotonin hypothesis is further favored by the fact that reserpine, a tranquilizer which contains an indole ring, depletes the brain of its store of serotonin and its tranquilizing effect parallels the time course of this depletion and not the time course of the presence of reserpine in the brain (Shore, Pletscher, Tomich, Carlsson, Kuntzman, & Brodie, 1957). It was deduced, therefore, that reserpine and its active analogues bring about tranquilization through interference with cellular binding of serotonin. Within the framework of the serotonin hypothesis, Brodie and Shore have proposed the following hypothetical scheme to account for the tranquilizing effect of reserpine and chlorpromazine and the hallucinogenic effects of LSD and mescaline. They propose that the hypothalamic parasympathetic centers are activated by serotonin-liberating, or "serotonergic" nerves, and the sympathetic centers by norepinephrine-

liberating, or adrenergic nerves. Dominance of the parasympathetic centers causes sedation or tranquilization, and that of the sympathetic centers, wakefulness. Reserpine, by preventing serotonin binding, will thus cause constant activation of the parasympathetic centers, while chlorpromazine, the other potent tranquilizer, would bring about sedation by blockade of adrenergic impulses to the central sympathetic centers and the consequent dominance of parasympathetic activity.

This paradigm permits Brodie and Shore to postulate the following mechanism of action of the two most potent hallucinogenic compounds. LSD blocks stimulation of the hypothalamic parasympathetic centers by interfering with the release of serotonin from the "serotonergic nerves," thus unmasking the action of the opposing sympathetic system. Mescaline, on the other hand, stimulates the "adrenergic" brain centers of the posterior hypothalamus directly, and mimics the action of norepinephrine. Thus an apparent parallelism between psychotic-like behavior and increased sympathetic activity produced by the hallucinogens has been proposed by Brodie and Shore, which tends in part to support Hoffer's emphasis upon the role of epinephrine and norepinephrine in the etiology of schizophrenia. A stricter interpretation of the Brodie-Shore hypotheses would restrict the effect of the tranquilizers and the hallucinogenic compounds to levels of activity, i.e., sedation, wakefulness, without any reference to psychotic behavior. For, at no time do the authors attempt to systematically relate the increased wakefulness resulting from excessive stimulation of adrenergic activity to psychotic or psychotic-like states. The only relationship that may be logically deduced from their writing is that psychotics are wide-awake.

Although the foregoing hypotheses regarding the role of serotonin in mental function seem attractive, there are some cogent criticisms which adherents of the serotonin paradigm have yet to answer. A close analogue of LSD, 2-bromo-D-lysergic acid diethylamide (Brom LSD) was found by Cerletti and Rothlin (1955) to be as effective as LSD in antagonizing several peripheral actions of serotonin, yet this antagonist of serotonin was found to be devoid of hallucinogenic action in humans. It should be noted that chlorpromazine is also a potent antiserotonin and instead of being hallucinogenic is a powerful antihallucinogen.

Before proceeding to a brief description of still another recent biochemical theory, it would serve well to re-emphasize how important a role the hallucinogens, specifically LSD and mescaline, have played in the two previous biochemical formulations of schizophrenia. The proponents of each theory have borrowed freely from each others' research efforts to substantiate hypotheses or propose biochemical nuances. This cross-fertilization implies that we have two distinct biochemical approaches which ultimately will coalesce into a unified biochemical theory of schizophrenia. This hope for the future is somewhat mitigated by a consideration of the clinical effects of LSD and mescaline which appear to be significantly different. Thus Matefi (1952) employing himself as a subject to study the effects of LSD and mescaline reported that they produced different psychopathologic reactions; the former, one of hebephrenic type and the latter catatonic. Drawings produced under the influence of LSD showed a tendency to expansion, while the "mescaline pictures" showed constriction.

Fischer, Georgi, and Weber (1951) also reported that LSD produced a predominantly hebephrenic state, whereas during mescaline intoxication catatonic features were outstanding. From such experiments in which subjects receive both drugs, it is a little difficult to see how presumably the same biochemical impairment can produce different experiences within the subject which are also significantly differentiable by an observer. These clinical observations lead to the obvious suggestion that the psychotic-like experience elicited by the hallucinogens is not homogeneous. Furthermore, as the behavioral concomitants of these toxic states are clearly differentiable, we may be dealing with two distinct biochemical processes, two distinct detoxification processes that are set off by various classes of chemical compounds. In short, considering the lack of *identity* between schizophrenia and the toxic states produced by sundry hallucinogens and the possibility of heterogeneity in mental dysfunction as specifically related to the class of compound administered to healthy people, the author may not have been considering biochemical theories of schizophrenia but rather biochemical theories of chemical psychoses. Future studies will undoubtedly clarify some of these issues.

Interference in Acetylcholine Metabolism

In addition to Hoffer, Pfeiffer and Jenney (1957) have emphasized the role of brain acetylcholine as a possible etiological factor in schizophrenia. In contrast to Hoffer who believes that schizophrenia is characterized in part by *excessive* central parasympathetic (cholinergic) activity, Pfeiffer and Jenney have presented evidence which supports the view that schizophrenics are *deficient* in acetylcholine. They base their argument on the facts that: (*a*) many tranquilizers have persistent acetylcholine-like effects; (*b*) drugs which show the nicotinic properties of acetylcholine, e.g., di-isopropyl fluorophosphate, make schizophrenics worse; and (*c*) drugs which show muscarinic effects, e.g., arecoline, make them better. In their study involving 23 schizophrenics they administered arecoline, a parasympathetic stimulant that passes the blood-brain barrier freely and leads to pronounced cholinergic effects, and methyl atropine nitrate, a quaternary analogue of atropine that protects against the peripheral effects of parasympathetic stimulants. Remarkable changes, it is claimed, ensued within 1–2 min. and lasted for 15 min. During a lucid interval which persisted for 15 min., the patients became more talkative, showed greater insight and sociability.

Recently, Rubin (1958a; 1958b) has reported the results of an experiment that may help to resolve the manifest contradiction between the theoretical formulations of Hoffer and Pfeiffer. It was shown that the hydrolysis rate of acetylcholine by human erythrocyte cholinesterase was significantly different between healthy people and psychotics. Furthermore, it was demonstrated that the psychotics were distributed between two discrete groups, each of which differed significantly from the normals with respect to the kinetics of enzymatic activity. One group hydrolyzed the substrate more rapidly than the normals while the other group hydrolyzed it more slowly. As rapid hydrolysis reduces the concentration of free acetylcholine, the former group may be assumed to be deficient in acetylcholine, and therefore corresponds to the defect postulated by Pfeiffer.

Defective enzymatic hydrolysis of acetylcholine in the latter group of patients may be characterized by excessive cholinergic activity, and they would deviate from Pfeiffer's model while congruent with Hoffer's hypothesis.

Summary

A brief review of several outstanding empirical studies suggests that schizophrenia is characterized by some disordered metabolic response to stress which in turn is dependent upon a neurohumoral or enzymatic defect. The three outstanding theories described have been derived from considerations of the chemical and pharmacological properties of the hallucinogenic or psychotomimetic agents. The fruitfulness of these theories and their biochemical constructs seems to be dependent upon the validity of the basic underlying equivalence relationship presumed to exist between the model psychosis produced by psychotomimetic agents and the pathological state(s) characterized as schizophrenia. The intense activity of researchers in this area should soon provide the relevant information required for evaluation of the basic assumption and possibly an effective chemotherapy for the schizophrenias.

REFERENCES

Ashby, W. Approach to solution of mental dysfunction through brain enzyme studies. *Amer. J. Psyhciat.*, 1950, **106**, 491–496.

Axelrod, J. O-methylation of epinephrine and other catechols in vitro and in vivo. *Science*, 1957, **126**, 400–401.

Axelrod, J. Is adrenochrome present in the blood? *Amer. J. Psychiat.*, 1958, **115**, 162–163.

Brodie, B. B., & Shore, P. A. A concept for a role of serotonin and morepinephrine as chemical mediators in the brain. *Ann. N. Y. Acad. Sci.*, 1957, **66**, 631–642.

Cerletti, A., & Rothlin, E. Role of 5-hydroxytryptamine in mental diseases and its antagonism to lysergic and derivates. *Nature, Lond.*, 1955, **176**, 785–786.

Fedoroff, S., & Hoffer, A. Toxicity of blood serum from schizophrenics and non-schizophrenic subjects. *J. nerv. ment. Dis.*, 1956, **124**, 396–398.

Fischer, R. Stress and toxicity of schizophrenic serum. *Science*, 1953, **118**, 409–411.

Fischer, R., Georgi, F., & Weber, R. Psychophysiche korrelationen VIII. Modellversuche zum schizophrenie problem. Lysergsäurediäthylamid und mezcalin. (Psychophysical relationships VIII. Experiments on the problem of schizophrenia with LSD and mescaline.) *Schweiz. med. Wschr.*, 1951, **81**, 817–837.

Hoffer, A. Epinephrine derivatives as potential schizophrenic factors. *J. clin. exp. Psychopath.*, 1957, **18**, 27–60.

Hoffer, A. Osmond, H., & Smythies, J. Schizophrenia: A new approach. II. Result of year's research. *J. ment. Sci.*, 1954. **100**, 29–45.

Holland, R., Cohen, G., Goldenberg, M., Sha, J., & Leifer, I. Adrenaline and noradrenaline in urine and plasma of schizophrenics. *Fed. Proc.*, 1958, **17**, 378.

Lea, A. J. Adrenochrome as the cause of schizophrenia: Investigation of some deductions from this hypothesis. *J. ment. Sci.*, 1955, **101**, 538–547.

Mátéfi, L. Mezealin—und Lysergsäurediäthylamid-Rausch. Selbstversuche mit besonderer Berücksichtigung eines Zeichentests. (Mescaline and lysergic acid diethylamide intoxication. Personal experiences with special consideration of a drawing test.) *Conf. Neurol.*, 1952, **12**, 146–177.

Örström, A. Isolation of phosphoglycolic acid from human erythrocytes. *Arch. Biochem.*, 1951, **33**, 484–485.

Osmond, H., & Smythies, M. B. Schizophrenia: A new approach. *J. ment. Sci.*, 1952, **98**, 309–315.

Pfeiffer, C. C., & Jenney, Elizabeth H. The inhibition of the conditioned response and the counteraction of schizophrenia by muscarinic stimulation of the brain. *Ann. N. Y. Acad. Sci.*, 1957, **66**, 753–764.

Reider, H. P. Biological determination of toxicity of pathologic body fluids. *Psychiat. Neurol.*, 1957, **134**, 378.

Resnick, O., Wolfe, J., Freeman, J., & Elmedilah, F. Metabolism of epineghrine in man. *Fed. Proc.*, 1958, **17**, 131.

RICHTER, D. (Ed.) Schizophrenia: Somatic aspects. New York: Macmillan, 1957.

RUBIN, L. S. Acetylcholine hydrolysis in psychiatric patients. *Science*, 1958, **128**, 254–255. (a)

RUBIN, L. S. Cholinesterase activity. *Science*, 1958, **128**, 1176–1178. (b)

SCHWARZ, B. E., WAKIM, K. G., BICKFORD, R. G., & LICHTENHELD, F. R. Behavioral and electroencephalographic effects of hallucinogenic drugs. *Arch. Neurol. Psychiat.*, 1956, **7**, 83–90.

SHAW, E., & WOOLLEY, D. W. Antiserotonins in hypertension and the antimetabolite approach to chemotherapy. *Science*, 1956, **124**, 121.

SHORE, P. A., PLECHTER, A., TOMICH, E., CARLSSON, A., KUNTZMAN, R., & BRODIE, B. B. Role of brain serotonin in reserpine action. *Ann. N. Y. Acad. Sci.*, 1957, **66**, 609–617.

WALASZEK, E. J., SMITH, C. M., & MINZ, B. Mechanism of inhibition of cortical reaction to adrenalin in rabbits treated with serum from schizophrenics. *Fed. Proc.*, 1958, **17**, 416.

WOOLLEY, D. W., & SHAW, E. A biochemical and pharmacological suggestion about certain mental disorders. *Science*, 1954, **119**, 587–588.

ZELLER, E. A. Biochemical and clinical changes resulting from administration of marsilid and tryptophan to schizophrenic patients. *J. clin. exp. Psycopath.*, 1958, **19**, 106–109.

Received February 2, 1959.

PSYCHOLOGICAL BULLETIN
VOL. 56, NO. 5, 1959

ON THE INTERACTION OF TWO SCALED VARIABLES

JEROME L. MYERS
University of Massachusetts

Statistical interaction has been most commonly defined as a measure of the joint effect of p variables upon performance. If a scaled independent variable is involved, the interaction sum of squares may be further analyzed, yielding more information than the above definition would indicate. This analysis was first presented by Alexander (1946), and subsequently extended by Grant (1956). This paper will present a complete analysis of the $(a-1)(b-1)$ degrees of freedom (df) involved in the interaction of two independent, scaled variables, A and B. The purpose is twofold: to extend the present state of comprehension of the interaction term and to introduce a method which will permit inferences about the rate of change of slope and curvature coefficients.

Grant has shown that if B is a scaled independent variable, with equal steps (or equal logarithmic steps), the $(b-1)(a-1)$ df can be analyzed into $b-1$ components, each on $a-1$ df. For example, if Ss in each of 3 groups ($a=3$) are tested on 5 consecutive days ($b=5$), one could compare the slopes of the curves for the three groups (linear component), the quadratic components of the three curves, the cubic components, and lastly the quartic components. Each of the four $(b-1)$ mean squares would be on two $(a-1)$ df. The appropriate error terms for these comparisons have also been presented by Grant.

If A and B are both scaled variables there are a number of alternative ways of analysing the data. We could, as above, compare the A curves over the levels of B. Alternatively, the B curves might be compared over the values of A. A more efficient analytical approach exists which will provide more information than can be obtained from performance of both the previously suggested analyses, and which will also shed light on the relation between the two ways of graphing the data.

The proposed analysis stems from one basic fact: if both variables are scaled, a sum of squares on 1 df can be computed for each of $(a-1)(b-1)$ components of the interaction sum of squares. Table 1 presents a set of data which has been analyzed in this manner. The design is a 5×3 factorial, with four entries in each of the 15 cells. The analysis may be approached in either of two ways. The B curves may be compared with respect to each of $a-1$ components; each sum of squares is then further analyzed into $b-1$ components. This has been done in the left half of Table 2. Alternatively, the A curves may be compared with respect to each of $b-1$ components, followed by further analysis of each sum of squares into $a-1$ components. This approach is presented in the right half of Table 2. The actual choice of approaches is of no consequence, since the same information can be obtained from either half of Table 2.

Consider the calculations for the left half of Table 2. The interaction sum of squares, computed in the usual manner, is 151.067. Further analysis requires reference to tabled values of orthogonal polynomials

TABLE 1

DATA AND COMPUTATIONAL AIDS FOR THE ANALYSIS

		A_1	A_2	A_3	$\sum_k^n Y_{ijk}$	$\sum_k^n X_1 Y_{ijk}$	$\sum_k^n X_1 Y_{ijk}$
B_1		20	18	16	54	− 4	0
		18	17	15	50	− 3	−1
		19	18	17	54	− 2	0
		16	16	15	47	− 1	−1
	B_1 Totals	73	69	63	205	−10	−2
B_2		18	15	14	47	− 4	2
		18	16	13	47	− 5	−1
		17	14	14	45	− 3	3
		16	13	13	42	− 3	3
	B_2 Totals	69	58	54	181	−15	7
B_3		16	12	11	39	− 5	3
		18	13	14	45	− 4	6
		17	12	13	42	− 4	6
		16	10	12	38	− 4	8
	B_3 Totals	67	47	50	164	−17	23
B_4		15	5	6	26	− 9	11
		18	8	8	34	−10	10
		17	7	9	33	− 8	12
		17	5	5	27	−12	12
	B_4 Totals	67	25	28	120	−39	45
B_5		17	7	6	30	−11	9
		18	10	9	37	− 9	7
		18	9	8	35	−10	8
		15	8	7	30	− 8	6
	B_5 Totals	68	34	30	132	−38	30
Totals		344	233	225	802	−119	103

(Fisher & Yates, 1953). For 4 *df* the values are:

$$\begin{array}{lccccc} Z_1 = & 2 & -1 & 0 & +1 & +2 \\ Z_2 = & +2 & -1 & -2 & -1 & +2 \\ Z_3 = & -1 & +2 & 0 & -2 & +1 \\ Z_4 = & +1 & -4 & +6 & -4 & +1 \end{array}$$

For 2 *df* the values are:

$$\begin{array}{lccc} X_1 = & -1 & 0 & +1 \\ X_2 = & +1 & -2 & +1 \end{array}$$

The subscript "1" refers to linear, "2" to quadratic, "3" to cubic, "4" to quartic. The sum of squares for the comparison of the linear component of the B curves may now be calculated.

$$SS_{B-linear} = \sum_j^b \left(\sum_i^n \sum_k^a X_1 Y_{ijk} \right)^2 \Big/ n \sum X_1^2 - \left(\sum_j^b \sum_i^n \sum_k^a X_1 Y_{ijk} \right)^2 \Big/ bn \sum X_1^2 \qquad [1]$$

where

i refers to scores within cells, $i = 1, 2, \cdots, n$.

j refers to levels of B, $j = 1, 2, \cdots, b$.

k refers to levels of A, $k = 1, 2, \cdots, a$. In our example $n = 4$, $a = 3$, and $b = 5$. The calculations are

$$SS_{B-linear} = [(-10)^2 + (-15)^2 + (-17)^2 + (-39)^2 + (-38)^2]/(4)(2) - (-119)^2/(5)(4)(2) = 93.350$$

This quantity is on 4 *df* and may now be analyzed into 4 components each on a single *df*. We have called these components *B-linear* (1), *B-linear* (2), etc. The calculations follow.

In our example

$$SS_{B-linear\ (1)} = [(-2)(-10) + (-1)(-15) + (0)(-17) + (1)(-39) + (2)(-38)]^2/(4)(10)(2) = 80.000$$

The $SS_{B-linear(2)}$ would be calculated by using value of Z_2, rather than Z_1, in Equation [2]. Thus

$$SS_{B-linear\ (2)} = [(2)(-10) + (-1)(-15) + (-2)(-17) + (-11)(-39) + (2)(-30)]^2/(4)(14)(2) = 0.571$$

Values of X_2, rather than X_1, are utilized in the calculations of $SS_{B-quadratic}$ and its components.

$$SS_{B-quadratic} = [(-2)^2 + (7)^2 + (23)^2 + (45)^2 + (30)^2]/(4)(6) - (103)^2/(5)(4)(6) = 57.717$$

$$SS_{B-quadratic\ (1)} = [(-2)(-1) + (-1)(7) + (0)(23) + (1)(45) + (2)(30)]^2/(4)(6)(10) = 43.350$$

By substituting values of Z_2, Z_3, and Z_4 for Z_1, $SS_{B-quadratic(2)}$, $SS_{B-quadratic(3)}$, and $SS_{B-quadratic(4)}$ can be computed.

$$SS_{B-linear\ (1)} = \left(\sum_j^b \sum_i^n \sum_k^a Z_1 X_1 Y_{ijk} \right)^2 \Big/ n \sum Z_1^2 \sum X_1^2 \qquad [2]$$

TABLE 2

TWO APPROACHES TO THE ANALYSIS OF THE INTERACTION

Source	*df*	Sum of Squares	Source	*df*	Sum of Squares
A×B	8	151.067	A×B	8	151.067
B-linear	4	93.350	A-linear	2	123.350
B-linear (1)	1	80.000	A-linear (1)	1	80.000
B-linear (2)	1	0.571	A-linear (2)	1	43.350
B-linear (3)	1	5.000	A-quadratic	2	5.821
B-linear (4)	1	7.779	A-quadratic (1)	1	0.571
B-quadratic	4	57.717	A-quadratic (2)	1	5.250
B-quadratic (1)	1	43.350	A-cubic	2	13.067
B-quadratic (2)	1	5.250	A-cubic (1)	1	5.000
B-quadratic (3)	1	8.067	A-cubic (2)	1	8.067
B-quadratic (4)	1	1.050	A-quartic	2	8.829
			A-quartic (1)	1	7.779
			A-quartic (2)	1	1.050

The mean squares on single *df*'s are measures of changes in slope and curvature of the *B* curves, over the levels of *B*. As an illustration, assume that we are interested in determining the function which describes the relation of dark adaptation *rate* (i.e. the slope of the dark adaptation curve) to preadapting light intensity. If we let *B* stand for intensity and *A* for trials, the $SS_{B-linear(1)}$ would enable us to test the hypothesis that rate of dark adaptation shows a linear decrement with increases in intensity. The $SS_{B-linear(2)}$ would enable us to test the hypothesis that the function relating rate of dark adaptation to intensity has a quadratic component. The $SS_{B-quadratic(1)}$ term is a measure of the extent to which the plot of the quadratic coefficient of the dark adaptation curve shows a linear change as a function of intensity. The Grant analysis yields a test of the hypothesis that the slopes (or quadratic, cubic, etc. components) of a number of curves do not differ. If they do differ, the analysis under discussion yields inferences about the way in which the component changes from curve to curve.

The analysis of the data in Table 1 was carried out by first computing the *B-linear* and *B-quadratic* components, then analyzing each of these into four components. We might have, as easily and meaningfully, computed four *A* components, then analyzed each of these into two com-

ponents. The results of these analyses are shown in Table 2. Note certain relationships between the two computational approaches. For example, *B-linear* (2) is equal to *A-quadratic* (1), *B-quadratic* (4) is equal to *A-quartic* (2). The *j*th component of the *k*th component of *A* will always equal the *k*th component of the *j*th component of *B*. Thus tests of hypotheses about the *B*-curves may be generated after an analysis of the *A*-curves by simply regrouping the single *df* components, and then adding. For example, *B-linear* is the sum of *A-linear* (1) and *A-quadratic* (1).

A discussion of error terms still remains. There are three possible cases: I. Each *S* is measured once under only one combination of *A* and *B*. II. There are different *S*s at each level of *B*, but all *S*s are measured at all levels of *A* (or *B* may be the within-*S*s variable, and *A* the between-*S*s variable). III. All *S*s are measured under all combinations of *A* and *B*.

Table 3 presents the error terms (denominator of *F*) for all interaction components, in each case.

Case I. The error term for all terms in Table 2 is the *within-cells* mean square. This value is 1.267. There are $ab(n-1)$ *df* or, in this case, 45. The *within-cells* sum of squares is, as always, the *total* sum of squares minus the *between-cells* sum of squares.

Case II. Assume that each row in Table 1 represents a different *S*. The *AB* interaction is tested against a $Ss \times A/B$ (subjects-by-*A*-within-*B*)

TABLE 3

ERROR TERMS FOR THREE DESIGNS

Case I		Case II		Case III	
Numerator of F	Denominator of F	Numerator of F	Denominator of F	Numerator of F	Denominator of F
All terms in Table 2	Within cells (1.267)	AB	$Ss \times A/B$ (.550)	AB	SAB (.538)
		A-linear		B-linear	error—B-linear (.738)
		A-quadratic		B-quadratic	error—B-quadratic (.340)
		A-cubic		A-linear	error—A-linear (.197)
		A-quartic		A-quadratic	error—A-quadratic (.903)
		B-linear and all its components	error-linear (.742)	A-cubic	error—A-cubic (.456)
				A-quartic	error—A-quartic (.600)
		A-linear (1)		B-linear (1) / A-linear (1)	error—B-linear (1) (.333)
		A-quadratic (1)			
		A-cubic (1)		B-linear (2) / A-quadratic (1)	error—B-linear (2) (1.000)
		A-quartic (1)		B-linear (3) / A-cubic (1)	error—B-linear (3) (.833)
		B-quadratic and all its components	error-quadratic (.358)	B-linear (4) / A-quartic (1)	error—B-linear (4) (.783)
		A-linear (2)		B-quadratic (1) / A-linear (2)	error—B-quadratic (1) (.061)
		A-quadratic (2)		B-quadratic (2) / A-quadratic (2)	error—B-quadratic (2) (.806)
		A-cubic (2)		B-quadratic (3) / A-cubic (2)	error—B-quadratic (3) (.078)
		A-quartic (2)		B-quadratic (4) / A-quartic (4)	error—B-quadratic (4) (.417)

error term on $b(n-1)(a-1)$ *df* (30 *df*). The computational formula is:

$$SS_{S\otimes A/B} = SS_{\text{Total}} - SS_{S_B} - SS_A - SS_{AB}$$
$$= 16.500 \qquad [3]$$

The mean square of .550 is the error term for *A-linear* (2 *df*), *A-quadratic*, *A-cubic*, and *A-quartic*, as well as for the *AB* interaction.

The linear component of error (*error-linear*) is computed from

$$SS_{error-linear} = \sum_j^b \sum_i^n \left(\sum_k^a X_1 Y_{ijk} \right)^2 \Big/ \sum X_1^2 - \sum_j^b \left(\sum_i^n \sum_k^a X_1 Y_{ijk} \right)^2 \Big/ n \sum X_1^2$$
$$= [(-4)^2 + (-3)^2 = \cdots + (-10)^2 + (-8)^2]/2$$
$$- [(-10)^2 (-15)^2 + \cdots + (-38)^2]/(4)(2)$$
$$= 11.125 \qquad [4]$$

The corresponding mean square ($df = b(n-1) = 15$) is .742. This is the error term for *B-linear* and all its components, and for *A-linear* (1), *A-quadratic* (1), *A-cubic* (1), and *A-quartic* (1).

The quadratic component of error (*error-quadratic*) is given by Equation [4] with values of X_2 substituted for X_1. The sum of squares is 5.375; the mean square is .358 on 15 *df*, and is the error term for *B*-quadratic and all its components, and for *A-linear* (2), *A-quadratic* (2), *A-cubic* (2), and *A-quartic* (2).

Case III. Assume that the first row at each level of *B* (Table 1) represents the performance of a single *S*, that the second rows represent a second *S*, etc. Thus we have four *S*s going through 15 combinations of *A* and *B*. Table 4 presents various sums of crossproducts which should facilitate calculations of the error terms for the single *df* components.

The error term for *AB* is *SAB*, and is computed in the usual manner. The resulting sum of squares is 12.933, which, when divided by 24 ($df = (n-1)(a-1)(b-1)$) yields a mean square of .538. The *SAB* sum of squares may next be analyzed into $a-1$ components, each on $(n-1)(b-1)$ *df*. In our example we have an *error-B-linear* and an *error-B-quadratic*, each on 12 *df*.

$$SS_{error-B-linear} = \sum_j^b \sum_i^n \left(\sum_k^a X_1 Y_{ijk} \right)^2 \Big/ \sum X_1^2 - \sum_j^b \left(\sum_i^n \sum_k^a X_1 Y_{ijk} \right)^2 \Big/ n \sum X_1^2$$
$$- \sum_i^n \left(\sum_j^b \sum_k^a X_1 Y_{ijk} \right)^2 \Big/ b \sum X_1^2 + \left(\sum_j^b \sum_i^n \sum_k^a X_1 Y_{ijk} \right)^2 \Big/ bn \sum X_1^2$$
$$= [(-4)^2 + (-3)^2 + (-2)^2 + (-1)^2 + \cdots + (-10)^2 + (-8)^2]/2$$
$$- [(-10)^2 + (-15)^2 + \cdots + (-38)^2]/(4)(2)$$
$$- [(-33)^2 + (-31)^2 + (-27)^2 + (-28)^2]/(5)(2) + (-119)^2/(5)(4)(2)$$
$$= 8.850 \qquad [5]$$

TABLE 4

COMPUTATIONAL AIDS FOR CASE III

Subject	Level of B	$Z_1 \sum_k^a X_1 Y_{ijk}$	$Z_2 \sum_k^a X_1 Y_{ijk}$	$Z_3 \sum_k^a X_1 Y_{ijk}$	$Z_4 \sum_k^a X_1 Y_{ijk}$	$Z_1 \sum_k^a X_2 Y_{ijk}$	$Z_2 \sum_k^a X_2 Y_{ijk}$	$Z_3 \sum_k^a X_2 Y_{ijk}$	$Z_4 \Sigma X_2 Y_{ijk}$
S_1	B_1	(−2)(−4)	(2)(−4)	(−1)(−4)	(1)(−4)	(−2)(0)	(2)(0)	(−1)(0)	(1)(0)
	B_2	(−1)(−4)	(−1)(−4)	(2)(−4)	(−4)(−4)	(−1)(2)	(−1)(2)	(2)(2)	(−4)(2)
	B_3	(0)(−5)	(−2)(−5)	(0)(−5)	(6)(−5)	(0)(3)	(−2)(3)	(0)(3)	(6)(3)
	B_4	(1)(−9)	(−1)(−9)	(−2)(−9)	(−4)(−9)	(1)(11)	(−1)(11)	(−2)(11)	(−4)(11)
	B_5	(2)(−11)	(2)(−11)	(1)(−11)	(1)(−11)	(2)(9)	(2)(9)	(1)(9)	(1)(9)
$\sum_j^b \sum_k^a ZXY_{1jk} =$		−19	−7	3	7	27	−1	−9	(−25)
S_2	B_1	(−2)(−3)	(2)(−3)	(−1)(−3)	(1)(−3)	(−2)(−1)	(2)(−1)	(−1)(−1)	(1)(−1)
	B_2	(−1)(−5)	(−1)(−5)	(2)(−5)	(−4)(−5)	(−1)(−1)	(−1)(−1)	(2)(−1)	(−4)(−1)
	B_3	(0)(−4)	(−2)(−4)	(0)(−4)	(6)(−4)	(0)(6)	(−2)(6)	(0)(6)	(6)(6)
	B_4	(1)(−10)	(−1)(−10)	(−2)(−10)	(−4)(−10)	(1)(10)	(−1)(10)	(−2)(10)	(−4)(10)
	B_5	(2)(−9)	(2)(−9)	(1)(−9)	(1)(−9)	(2)(7)	(2)(7)	(1)(7)	(1)(7)
$\sum_j^b \sum_k^a ZXY_{2jk}$		−17	−1	4	24	27	−9	−14	6
S_3	B_1	(−2)(−2)	(2)(−2)	(−1)(−2)	(1)(−2)	(−2)(0)	(2)(0)	(−1)(0)	(1)(0)
	B_2	(−1)(−3)	(−1)(−3)	(2)(−3)	(−4)(−3)	(−1)(3)	(−1)(3)	(2)(3)	(−4)(3)
	B_3	(0)(−4)	(−2)(−4)	(0)(−4)	(6)(−4)	(0)(6)	(−2)(6)	(0)(6)	(6)(6)
	B_4	(1)(−8)	(−1)(−8)	(−2)(−8)	(−4)(−8)	(1)(12)	(−1)(12)	(−2)(12)	(−4)(12)
	B_5	(2)(−10)	(2)(−10)	(1)(−10)	(1)(−10)	(2)(8)	(2)(8)	(1)(8)	(1)(8)
$\sum_j^b \sum_k^a ZXY_{3jk} =$		−21	−5	2	8	25	−11	−10	−16
S_4	B_1	(−2)(−1)	(2)(−1)	(−1)(−1)	(1)(−1)	(−2)(−1)	(2)(−1)	(−1)(−1)	(1)(−1)
	B_2	(−1)(−3)	(−1)(−3)	(2)(−3)	(−4)(−3)	(−1)(3)	(−1)(3)	(2)(3)	(−4)(3)
	B_3	(0)(−4)	(−2)(−4)	(0)(−4)	(6)(−4)	(0)(8)	(−2)(8)	(0)(8)	(6)(8)
	B_4	(1)(−12)	(−1)(−22)	(−2)(−12)	(−4)(−12)	(1)(12)	(−1)(12)	(−2)(12)	(−4)(12)
	B_5	(2)(−8)	(2)(−8)	(1)(−8)	(1)(−8)	(2)(6)	(2)(6)	(1)(6)	(1)(6)
$\sum_j^b \sum_k^a ZXY_{4jk} =$		−23	5	11	27	23	−12	−11	−7

The mean square is .738 and is the error term for *B-linear*. The sum of squares for the *error-B-quadratic* is obtained by substituting X_2 values in Equation [5]. The resulting sum of squares is 4.083, and the mean square is .340, the latter being the error term for *B-quadratic*. The *SAB* term could also have been analyzed into $b-1$ components, each on $(n-1)(a-1)$ *df*. Equation [5] would then contain Z values rather than X values, and would yield the sums of squares for *error-A-linear*, *error-A-quadratic*, etc. These in turn would provide F tests for *A-linear*, *A-quadratic*, etc.

Next consider the error terms for *B-linear* and *B-quadratic* components. The sum of squares for *error-B-linear* (1) is given by

$$SS_{error-B-linear\ (1)} = \sum_i^n \left(\sum_j^b \sum_k^a Z_1 X_i Y_{ijk} \right)^2 \Big/ \sum X_1^2 \sum Z_1^2 - SS_{B-linear\ (1)}$$
$$= [(-19)^2 + \cdots + (-23)^2]/(2)(10) - 80.000 \qquad [6]$$
$$= 1.00$$

The mean square on $n-1$ $(=3)$ *df* is .333 and is the error term for *B-linear* (1). The remaining error components (mean squares) are found in Table 3, and were calculated in similar fashion, by substituting appropriate values of X and Z in Equation [6]. Relationships previously pointed out with respect to the *AB* interaction hold for *SAB*. Thus, *error-B-linear* (2) is identical to *error-A-quadratic* (1), for example. This indicates that *error-A-linear*, *error-A-quadratic*, etc. can be obtained by adding appropriate *error-B* components. For example, *error-B-linear* (1) and *error-B-quadratic* (1) should equal *error-A-linear*.

The analyses described permit inferences about such matters as the rate of learning as a function of amount of practice, the rate of dark adaptation as a function of pre-adaptation intensity, or the rate of extinction as a function of number of conditioning trials. While such information may appear meaningful and useful, the reader may wonder if tests of quantities such as *B-quadratic* (4) are of any utility. To this, it may be answered that the complete single *df* analysis should extend our understanding of the interaction term. Furthermore, calculation of additional terms provides computational checks for those components which are of major interest. Finally, it is hoped that the information obtained from such analyses will form some basis for more quantitative, specific, and accurate predictions about behavior than our present theories generally yield.

REFERENCES

ALEXANDER, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.

FISHER, R. A., & YATES, F. *Statistical tables for biological, agricultural, and medical research* (4th Ed.) New York: Hafner, 1953.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

Received January 12, 1959.

PSYCHOLOGICAL BULLETIN
VOL. 56, NO. 5, 1959

MULTIPLE COMPARISONS IN ANALYSIS OF VARIANCE

JOHN GAITO
Wilkes College

Recently Ryan (1959) has provided a valuable service for psychologists by presenting newer techniques for making multiple comparisons after the analysis of variance *F* test has rejected the over-all null hypothesis that the means are equal. Some of these techniques previously were not available to psychologists because they were discussed in scattered sources. However, Ryan does not mention the fact that the analysis of variance technique does provide a means of making individual comparisons (after rejection of the null hypothesis) which is suitable for many situations. This procedure involves partitioning the *n* degrees of freedom for the main effect into *n* orthogonal components, each with one *df* (Edwards, 1951; Senders, 1958; Snedecor, 1946) and is suitable for either one variable or multiple variable designs.

As an example of this procedure let us take the case in which we have three groups of *S*s. Group 1 is a control; Groups 2 and 3 are experimental groups. Inasmuch as three groups are involved, two *df* are present. Therefore, we can partition the sum of squares with two *df* into two single components, each with one *df*. The meaningful comparisons in this situation would be G_1 vs. G_2+G_3, and G_2 vs. G_3.

As another example let us have two experimental groups (E_1 and E_2) and a control group for each (C_1 and C_2). Here we have a sum of squares for between groups with three *df*. We partition these three *df* into three single components, each with one *df*. The experimental hypotheses would determine which comparisons would be relevant. In this case it appears that the investigator would be concerned with a comparison between (C_1+E_1) and (C_2+E_2) for one *df*. A second *df* would involve the comparison, C_1 vs. E_1; the third *df*, C_2 vs. E_2. Each of these three tests would be by the *F* ratio, the denominator being that which is provided by the analysis of variance model and would be the same for all three tests. This procedure allows one to make use of the power of the variance analysis technique (a well-explored robust procedure) rather than to develop new procedures. Thus the probability statements would be exact and the error rate per comparison would be that of the probability level used for the tests of significance, as is usual with the analysis of variance procedure. This would maintain the error rate per experiment at a lower rate than would occur if all possible comparisons were effected.

With the partitioning procedure it is important that the *orthogonal comparisons be planned before the experiment is conducted*. In the second example, if the investigator had not planned the comparisons, he might be tempted to make different orthogonal comparisons than those above after he looks at the data. For example, he might compare (E_1+E_2) and (C_1+C_2), E_1 vs. E_2, and C_1 vs. C_2. However, which orthogonal comparisons are to be effected will be determined by the experimental hypotheses. (For a simple means of determining orthogonality, see Senders, 1958.)

Using this procedure, Ryan's Case

1 (multiple comparisons) reduces essentially to his Case 3 (multiple variables in analysis of variance). A single variable might be involved but the partitioning into n components gives the analysis the appearance of a multiple variable design, according to procedure in making tests of significance. Thus, two or more F ratios will be involved and, in a strict sense, these ratios will not be independent because a common error estimate will be used. However, this lack of independence is irrelevant. The important point is that the numerator and denominator of the F ratio be independent. Each F ratio is based on the mathematical model (the F distribution) and has all the power of that model.

The above examples indicate the suitability of the partitioning procedure for many cases. However, in some experiments the partitioning procedure will not be completely efficacious inasmuch as all comparisons which are required by the experimental hypotheses may not be orthogonal. Yet if most comparisons are orthogonal ones it would appear that little loss of exactitude would occur with few nonorthogonal comparisons. Sometimes the results of the orthogonal comparisons will provide indirect information concerning the relative ranking of all different groups, thus obviating the need for making nonorthogonal comparisons. But in this situation the new multiple comparison techniques which Ryan discusses might be used. However, it should be pointed out that a trend analysis (Lindquist, 1953) might be more appropriate if the groups differ quantitatively (or qualitatively if the groups can be ordered). However, a regression analysis may be more meaningful if the groups differ in an orderly quantitative manner.

REFERENCES

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1951.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

RYAN, T. A. Multiple comparisons in psychological research. *Psychol. Bull.*, 1959, **56**, 26–47.

SENDERS, V. L. *Measurement and statistics.* New York: Oxford, 1958.

SNEDECOR, G. W. *Statistical methods.* Ames: Iowa State Coll. Press, 1946.

Received February 12, 1959.

PSYCHOLOGICAL BULLETIN
VOL. 56, NO. 5, 1959

COMMENTS ON ORTHOGONAL COMPONENTS

T. A. RYAN
Cornell University

Professor Gaito is to be thanked for pointing out a method for making multiple comparisons which was not discussed in my paper. Long as that paper was, it still could not cover all of the possible approaches. The method of orthogonal components which Gaito proposes has, however, several serious drawbacks. Moreover, the power of the method is little, if any, greater than that of other methods which are not subject to these difficulties.

The following points must be kept in view in evaluating the method advanced by Gaito:

1. By restricting ourselves to orthogonal comparisons, we are prevented from making comparisons which may be important in understanding and interpreting the results. In his first example involving two experimental groups and a single control, Gaito asks two questions: (*a*) whether the two experimental groups differ jointly from the control group, and (*b*) whether the experimental groups differ from each other. If we ask these two questions we cannot then ask *which* of the experimental groups differ from the control.

2. When the number of groups is larger than that in Gaito's example, we are even more restricted in the questions which we can ask. For example, if there are 5 experimental groups and 1 control, we have a total of 5 degrees of freedom and therefore 5 orthogonal comparisons are possible. There are only 4 degrees of freedom for question (*b*), so we can make only 4 of the 10 possible comparisons among the experimental groups.

Dunnett (1955) has provided a method for comparing each experimental group with a control while controlling the error rate experimentwise. This method makes the same number of comparisons as the method of orthogonal components, but the comparisons it does allow are more meaningful for many common situations. If we use Dunnett's method we cannot find out which of the experimental groups differ from each other. To ask these questions brings us to the situation where we make all possible comparisons among the means, and a method like Tukey's is the most appropriate.

3. The relative power of two methods of comparison can be determined adequately only if the error rates are computed on the same basis. Dunnett's and Tukey's methods are designed to control the error rate *experimentwise*. To compare these with the method of orthogonal comparisons it is necessary to determine the experimentwise error rate for the latter. When this is done, as has been illustrated in Table 1, we find that the Dunnett method actually requires slightly smaller differences for significance than does the method proposed by Gaito.[1] (We must bear in mind, of course, that the Dunnett method is not comparing the same

[1] In order to determine allowances for the method of orthogonal components in Table 1, the F (or t) value was determined at the tabled ".025 level." Since there are two comparisons in the example, the error rate *per experiment* is twice the nominal level of the single tests, or .05. The *experimentwise* rate is only slightly less (.04994).

TABLE 1

COMPARISON OF ALLOWANCES FOR 5% LEVEL EXPERIMENTWISE OR PER EXPERIMENT, TWO EXPERIMENTAL GROUPS, AND ONE CONTROL

df for Error	F^* for Orthogonal Components	$t=\sqrt{F}$	Critical t for Dunnett Method**	Critical t for Tukey Method***	
				Extremes	Adjacent Means
15	6.20	2.49	2.46	2.62	2.41
30	5.57	2.36	2.32	2.47	2.29
∞	5.02	2.24	2.21	2.35	2.17

* Read from F tables for ".025 level" which gives .05 level *per* experiment, with two comparisons.
** Allowance for comparing each of two experimental means with control mean, at .05 level *experimentwise*.
*** For all comparisons at .05 level experimentwise, computed from Tukey's tables.

pairs as the method of orthogonal comparisons, so we are comparing the power of one in making *its* comparisons with the power of the other in making its own.) Tukey's method, which permits us to make all possible comparisons of pairs of means, requires a somewhat greater gap when we are comparing the extreme means, but a smaller separation for adjacent pairs.

4. When it has been decided in advance that only certain comparisons are of interest, and that other comparisons will never be made under any circumstances, multiple comparison procedures can be adapted to control the error rate *per experiment*. To do so, it is not necessary to ensure that all comparisons are orthogonal. The error rate is simply the product of the number of comparisons to be made and the nominal significance level of the individual tests (the error rate per comparison). This relation is not affected by the lack of independence of the comparisons (Ryan, 1959, p. 39).

Allowances based upon *experimentwise* rates may be somewhat smaller than those for the corresponding rates *per experiment*. This is the reason for the difference between Dunnett's allowances and those based upon F in Column 3 of Table 1. The allowances for the Tukey method would have been slightly larger if we had computed them on the basis of rates per experiment. For example, the allowance for the extremes in the case of 15 degrees of freedom would have been 2.73 instead of 2.62, but this is the largest change which would have been made in the table. When we select our comparisons, however, a method which controls the experimentwise rate is usually not readily available. What we may lose by the shift to the rate per experiment may nevertheless be compensated for by the reduced number of comparisons.

5. In describing his procedure, Gaito has used the error rate per comparison. This is also the rate per degree of freedom since the number of comparisons equals the degrees of freedom. In my paper, however, I argued that we should control the rate of error per experiment or experimentwise. I shall not review these arguments here, since Gaito has not offered any reasons for preferring the rate per degree of freedom.

Even if we were convinced that the error rate should be based upon the number of degrees of freedom, there are other more satisfactory methods available. Duncan's meth-

ods (1955) permit us to make all possible comparisons with an error rate which is proportional to the number of degrees of freedom. The Duncan test will require slightly greater separation of a pair of means for significance, in comparison to those in Gaito's method, but this loss of power is more than compensated for by the possibility of asking more meaningful questions in our analysis of the data.

Apart from this analysis of the method proposed by Gaito, I wish to correct a misunderstanding which may arise from his first sentence. The newer methods for multiple comparisons do *not* require an initial *F* test of the over-all variation between groups. All of the methods mentioned in the present note can be applied immediately, without such a preliminary test.

Having raised a number of points of disagreement, I can nevertheless end with an important area of agreement, namely the possibility of using regression or trend analysis in cases where the groups can be ordered or classified according to a quantitative independent variable. This whole approach was not considered in my paper because of limitations of space and because it is generally much better understood than the cases of qualitative groupings. At least, in regression analysis, the number of significance tests does not increase geometrically as the number of groups or subdivisions goes up.

REFERENCES

DUNCAN, D. B. Multiple range and multiple F tests. *Biometrics*, 1955, 11, 1–42.

DUNNETT, C. W. A multiple comparison procedure for comparing several treatments with a control. *J. Amer. statist. Ass.*, 1955, 50, 1096–1121.

GAITO, J. Multiple comparisons in analysis of variance. *Psychol. Bull.*, 1959, 56, 392–393.

RYAN, T. A. Multiple comparisons in psychological research. *Psychol. Bull.*, 1959, 56, 26–47.

Received March 2, 1959.

ERRATUM

In "Multiple Comparisons in Psychological Research," *Psych. Bull.*, Vol. 56, No. 1, pp. 26–47, there is an error in the next to last sentence of the appendix. This sentence should read:

"Before M_3 can differ significantly from M_4, *both* M_3 must be found to be significantly different from M_5, *and* M_4 must differ significantly from M_2."

In other words, a given pair of means cannot differ significantly in the layer method unless *all* larger subgroups to which it belongs are also significant. When a particular subgroup is found to have a nonsignificant range, no further tests are made within it, just as the total range must be significant for any further tests. (This point is correctly stated on page 41.)